DOCUMENT RESUME

ED 305 059	IR 013 735
TITLE	Miscellaneous Papers Presented at the Telecommunications Policy Research Conference.
INSTITUTION	Telecommunications Policy Research Conference, Inc., Washington, DC.
PUB DATE	Oct 88
NOTE	963p.; Papers presented at the Telecommunications Policy Research Conference (16th, Airlie House, Warrenton, VA, October 30-November 1, 1988). For other papers of this conference, see IR 013 715-734.
PUB TYPE	Speeches/Conference Papers (150)
EDRS PRICE	MF07/PC39 Plus Postage.
DESCRIPTORS	Cable Television; Competition; *Economics; *Federal Regulation; Networks; *Policy Formation; Radio; Research; Technology; *Telecommunications; *Telephone Communications Industry; Television; Utilities; Video Equipment
IDENTIFIERS	*Divestment; Pricing; Rate of Regulation; *State Regulation

ABSTRACT

This document consists of 26 miscellaneous conference papers that consider various aspects of telecommunications policy, as follows: (1) "New Communications Systems and Services" (Baer); (2) "Regulatory Incentive Mechanisms and Political Action" (Baron); (3) "Efficiency Issues in Competitive Allocation of the Broadcast Spectrum" (Borenstein); (4) "Divestiture Policy Considerations in an Information Services World" (Brennan); (5) "The Effects of Household Characteristics on Telephone Usage and Class of Service Choice" (Cox); (6) "Modeling the Effects of Household Characteristics on Telephone Usage and Class of Service Choice" (Cox and Ruud); (7) "Regulation Games" (Gilbert and Newbery); (8) "Exploring the Hidden Costs of Overbuilds" (Glist); (9) "The Economics of the 'New' Hollywood" (Gomery); (10) "Privatization of NTT in Liberalized Japanese Telecommunications Market" (Hayashi and Toda); (11) "Repeated Auctions of Incentive Contracts, Investment, and Bidding Parity with an Application to Takeovers" (Laffont and Tirole); (12) "The Economics and Pricing of Broadband" (Levin); (13) "Ratings Policy and Public Policy" (Miller); (14) "Open Interconnection and the Economics of Networks" (Mueller); (15) "The Political Economics of State Responses to Divestiture and Federal Deregulation in Telecommunications" (Noll and Smart); (16) "The Uniform Settlements Policy in International Telecommunications" (O'Brien); (17) "The Conflict Between Spectrum Efficiency and Economic Efficiency" (Rogers and Woodbury); (18) "Trigger Price Regulation" (Salant and Woroch); (19) "Behind the Scenes Power in Japan's Information Industry" (Salvaggio); (20) "Alternatives to Rate of Return Regulation in the Telecommunications Industry" (Sievers); (21) "State Regulators, Legislators, and U S West" (Teske); (22) "Frequency Spectrum Deregulation, Property Rights and Markets" (Webbink); (23) "ATV Standards and Trade in Recorded Video Entertainment" (Wildman); (24) "Estimation of a Market Growth Model with Network Externalities and Dynamic Pricing Using an Extended Kalman Filter" (Xie and Sirbu); (25) "Reneging by Cable Operators on Their Franchise Promises" (Zupan); and (26) "Cable Franchise Renewals" (Zupan). The texts are supplemented by various tables and figures; most of the papers also contain references. (EW)

ED305059

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

MISCELLANEOUS PAPERS

Presented at

TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE

Arlie Virginia

October 30-November 1, 1988

IR013735

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gerald R. Faulhaber

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

September 26, 1988
Draft #3

NEW COMMUNICATIONS SYSTEMS AND SERVICES:

TECHNOLOGICAL AND POLICY CHOICES

Walter S. Baer

Times Mirror Company
Times Mirror Square
Los Angeles, CA 90053
(213) 237-3781

This draft is for review and comment
only. It is not to be quoted or cited.

IR013735

CONTENTS

I TECHNOLOGY AND POLICY INTERACTIONS 1
Technical Trends in Transmission and Processing 2
Transmission and Processing Tradeoffs 5
Adoption of Communication Technologies 9
Regulatory Influences on Innovation 14
Technology's Impact on Policy 16

II HIGH DEFINITION TELEVISION 20
Japan's HDTV Developments 20
Alternative Systems for the United States 22
Setting HDTV Standards 30

III BROADBAND DISTRIBUTION NETWORKS 33
Telco Scenarios 33
Cable Scenarios 37
Policy Issues 39

IV TECHNOLOGY AND POLICY CHOICES 43

W. S. Baer
Draft #3
September 26, 1988

New Communications Systems and Services: Technological and Policy Choices

I. TECHNOLOGY AND POLICY INTERACTIONS

Technology will be the principal driving force behind the introduction of new communications systems and services in the 1990s. As examples, more powerful communications satellites, fiber optic transmission links, digital switches, advanced television systems, and electronic information services are under development in many parts of the world. Technology is often considered an independent, exogenous force whose pace cannot be predicted; but like other human endeavors, technology develops in a particular social and political context. Regulatory, economic, and other public policies influence technology just as technology impacts policy.

This paper examines technological trends and developments expected over the next decade and explores their likely interaction with communications policy in the United States. The focus is on new communications products and services, with particular emphasis on consumer rather than business applications.[1] High definition television (HDTV) and broadband distribution networks are discussed in greater depth, both because they appear high on policymakers' agendas, and because they illustrate the interactions between technology and policy.

Technical Trends in Transmission and Processing

The technologies driving new communications and information services continue to advance, with no early saturation in sight. "More, better, and cheaper" characterize the technical trends for both transmitting and processing information, introducing new tradeoffs in system design and utilization.[2]

Among transmission technologies, digital fiber optic systems show the most spectacular gains. The fiber pairs installed in Los Angeles during the 1984 Olympics carried 45 megabits per second (45 Mbps), enough capacity for 672 digital voice channels. Fiber systems installed in 1988 can carry 565 Mbps, a twelve-fold increase in four years. They have become the transmission medium of choice for nearly all new links between telephone central offices. Fiber optic systems planned for the 1990s will have capacities in the thousands of megabits. And although fiber optics are advancing most rapidly, satellite, coaxial cable, microwave and other terrestrial transmission systems are improving as well. They present moving targets for performance/cost comparisons with fiber.

At the same time, the processing of information is becoming faster and cheaper. The pace of development is particularly striking in the underlying technologies of semiconductors, optoelectronics, and computing:

- Semiconductor very-large-scale integration (VLSI) largely follows the path predicted over a decade ago,[3] with more than a fifty-fold improvement in the state-of-the-art in ten years. Silicon VLSI circuits now contain more than 1 million logical components (gates) on a single chip, with more computing power than most computer mainframes of the 1970s.

- New techniques of chip design ("silicon compilers") enable low cost production of communications components such as digital coders/decoders (codecs) and signal processors.

- Although VLSI chips today are made from silicon, gallium arsenide and related compounds promise even higher speed performance for digital switches and integrated optic/electronic devices.

- Parallel processing computer architectures are becoming commercially available and should find important applications in telecommunications switching and information retrieval systems.

- New discoveries in superconductivity may result in even faster, more efficient switches and processors in the long run, but superconductivity does not seem likely to play a significant role in communications systems in the next ten years.

These developments feed directly into improved telecommunications switches, PABXs, facsimile scanners, cellular telephones, and virtually all other communications equipment. At the consumer end, one result of the advances in integrated circuits will be the introduction of a wristwatch pager in early 1989.

Software engineering also is advancing, although not so rapidly as hardware technologies. Software development dominates the cost of most new communications systems; more than half of the technical staff at AT&T Bell Laboratories and Bell Communications Research work on software problems. Software trends include distributing intelligence throughout complex networks, and developing "expert systems" for routine network operation, maintanance and control.[4] Software improvements also make it easier to interact with communications or information systems. For example, voice response systems let callers use push-button telephones to access computer-based information such as stock quotes, airline schedules and theater ticket offerings. Software that could recognize human voice commands would be much more powerful, but the technology of general-purpose voice recognition is still not commercially available. Developers hope that voice recognition software will be integrated in communications systems sometime in the next decade.

The technical advances in information transmission and processing accelerate the trend to digital communications which began in the 1960s. Although the U.S. telephone network remains a mixture of analog and digital facilities, it continues to evolve in the digital direction toward the planned Integrated Services Digital Network (ISDN). As currently conceived, the ISDN will provide customers with end-to-end digital service at 144 kbps -- specifically, two 64 kbps circuit-switched channels plus a 16 kbps packet-switched channel. Although implementing this "narrowband" ISDN will take years and billions of investment dollars, it can utilize existing copper wires, and the 64 kbps switching technology it requires is commercially available today.[5]

However, the narrowband ISDN concept may already be too restrictive for some business customers who want higher speed data and video capacities.[6] International standards groups are now working on a "broadband" ISDN that would provide end-to-end digital communications at speeds of 150 Mbps or more -- 1000 times greater than the narrowband ISDN.[7] Fiber optic transmission systems can readily handle these data rates, but high speed switching in the megabit-per-second range is still a laboratory phenomenon. Broadband ISDN services presumably would be implemented first to commercial customers with dedicated fiber optic links. Switched broadband ISDN services to businesses and homes remain a telco vision for the late 1990s or beyond (see Section III).

Transmission and Processing Tradeoffs

Communications system designers have always had to consider tradeoffs between transmission capacity and information processing at the sender's and receiver's ends. Telegraph lines in the 19th Century had little information carrying capacity, so that operators encoded messages into a series of short and long pulses (dots and dashes) and decoded them at the receiving end. Today, computers process information at rates measured in millions-of-instructions-per-second, and transmission lines operate at hundreds of megabits per second, but tradeoffs still are necessary.

Facsimile transmission provides a good illustration. Facsimile (fax) requires equipment to scan a document and convert the information to analog or digital signals, a transmission link from sender to receiver, and equipment at the receiving end to process the information and print it on paper. Before 1980, facsimile scanners produced analog signals for transmission over analog telephone lines. Sending a page of text took from three to six minutes. Advances on the processing side led to digital fax machines that transmit a page over the same standard telephone lines in less than one minute. The next advance has been to use digital transmission lines with higher data rates, requiring only a few seconds to transmit a fax page. Fax transmission speed in the past decade thus has increased by about a factor of fifty, both from improvements in fax signal processing and from the greater transmission capacity available on new digital lines. The cost of facsimile equipment has also decreased dramatically, increasing the number of fax users and opening up new markets.

Another example of information processing/transmission tradeoffs is the growing use of local information retrieval systems, based on optical storage devices and more powerful personal computers. In the past, large databases were expensive to store and replicate. Consequently, users accessed them at a central site over telephone lines and incurred both communications and processing charges with each inquiry. Development of CD-ROM and related optical devices capable of storing hundreds of megabytes[8] can now bring these databases to the user for local searching on his or her PC. The cost of the database and its updates will generally be fixed, with no additional charges per inquiry. As a result, frequent database users will increasingly substitute local retrieval for long distance access of remote databases in the next decade.

Video transmission provides a third illustration. Transmitting a standard television signal in the United States requires a 6 megahertz (6MHz) analog bandwidth. Using straightforward digitizing techniques, the same signal requires a digital transmission rate of 90 megabits per second (90 Mbps). However, processing the picture before transmitting it can reduce the effective bandwidth or bit rate required. Bandwidth compression equipment has been available for several years to process analog television signals in a way that two television programs can be transmitted over a single 6MHz channel.[9] The cost of the processing equipment at either end must be compared with the cost of an extra transmission channel. Generally, for coaxial cable systems that have large channel capacity, "two for one" analog signal processing has not been worth the cost. This approach might be economic, however, for some microwave or satellite transmission applications where bandwidth is at a premium.

Digital processing of television signals, using data compression techniques that remove redundancies in the signal, can reduce the transmission bit rate by a factor of two-to-four with no loss in perceived picture quality. Further compression requires some compromise in the amount and quality of motion displayed. Slowly moving pictures adequate for most teleconferences can be compressed to 1.5 Mbps or even 64 kbps, data rates that are now becoming widely available throughout the U.S. telephone network, but these compression schemes are not suitable for the degree of motion required for entertainment television. Still pictures can be readily compressed and transmitted over standard telephone lines in under a minute. Japanese manufacturers have developed such "slow scan" terminals, selling for under $500, for consumer as well as business markets.

The trade-offs between communications transmission and processing are dynamic ones. There is no final answer. In his 1987 Report on Competition in the Telephone Industry, Peter Huber emphasized the technical trends favoring communications processing and switching:

> "In the past hundred years, the basic low-density transmission technology--twisted copper wire routed through underground conduits or via overhead poles--has not changed much, nor has its price. Fiber-optic cable has slashed transmission costs for high-density applications, but the last mile of the network, where about half the transmission expense arises, carries mostly low-density traffic. In the past fifteen years, on the other hand, revolutionary developments in electronics have slashed the costs of

> switching and other forms of network intelligence. The inexorable trend is therefore to move switching out toward the end user. Making connections at a distance costs more in increased transmission then it saves in lowering the number of connections needed."[10]

These trends have indeed resulted in distributing intelligence within networks and substituting processing for transmission. The rise of packet switching provides one illustration. But advances in transmission technologies may change the balance again. With the carrying capacities of fiber optic (and other) transmission systems increasing so rapidly, packet switched services may lose some of their appeal. Similarly, we may see a glut of long haul transmission capacity in the next several years that will bring more price competition and accelerate the shakeout among long distance carriers.

Adoption of Communication Technologies

Turning to the communications end-user, we can see how technology's advance has brought new products and services to the marketplace over the past decade. The first large-scale test of cellular mobile telephone service was begun in Chicago in 1978. Commercial service started in 1983, and now, only five years later, more than 180 cellular systems are in operation serving more than one million subscribers.

Cable television's reach has expanded from fewer than 20% of U.S. households in 1978 to more than 50% today. The penetration of videocassette recorders (VCRs) has grown even more rapidly than cable. VCRs are now found in well over half of U.S. households. For many families, watching a videotape has replaced Saturday night television viewing and moviegoing. Videotape sales and rentals have, in fact, become the largest source of revenue for motion pictures, surpassing theater box office receipts.[11]

The personal computer (PC) has changed from a curiosity ten years ago to a fixture in offices and a significant number of homes. Although personal computers themselves are not yet part of the normal home routine, microprocessors have been integrated into a wide range of consumer products. Telephone answering machines, VCRs, television receivers, and addressable cable converters all contain sophisticated microprocessors that control their increasingly complex functions. Microprocessors are also widely used as control devices in refrigerators, microwave ovens, washers, dryers, and sewing machines. Automobiles now incorporate microprocessors for control functions from adjusting the correct fuel mixture to remembering the proper seat adjustment for each driver. And during the 1980s, consumers have become accustomed to using computers disguised as automatic teller machines (ATMs), in-store ordering kiosks, and pay telephones.

Just as Moliere's bourgeois gentleman expressed surprise at his ability to speak prose, consumers might now be astonished by the number and variety of microprocessors they use in everyday life. The situation is analogous to the development and diffusion of electric motors. Sixty years ago, electric motors were too large and too expensive for most consumer applications. The development of cheap, fractional-horsepower motors led to the proliferation of small motorized appliances after World War II. Today, American consumers typically own dozens of electric motors in power tools, can openers, knives, mixers, blenders, toys and toothbrushes. In similar fashion, microprocessors are proliferating, quietly and unseen, throughout the consumer economy.

A look back at these developments over the past decade suggests some general points about how consumers adopt new communications technologies. First, changes occur incrementally over a period of years after a new technology or service has been introduced. Even color television, which nearly all television viewers wanted when they saw it, took more than 10 years to reach 50% of U.S. households (see Table 1). Cable television took four decades to achieve 50% household penetration. The most rapid adoption occurs for new products, like color TV sets or VCRs, that can be purchased as individual units. Technologies that must be integrated into networks, like digital PBXs, packet switching, or fiber optic distribution systems, take much longer to diffuse.

Second, consumer adoption of new technologies and services typically follows a top-down pattern. More affluent, better educated consumers are the first to try a new technology or service -- not only because they have more money to spend, but because their life styles prompt them to be early adopters.[12] The rest of the market follows in the predictable "s-curve" pattern shown in diffusion studies. Compact disc audio players and telephone answering machines seem to be following that diffusion curve toward mass market acceptance. But other new products -- like home banking and videotex -- have yet to show much appeal beyond a limited, elite audience.

Many communications innovations also meet with early consumer resistance, even when they find wide acceptance later. Telephone answering machines provide a good illustration. When introduced tothe consumer market in the early 1970s, many people found them "inhuman," "insulting," and "intimidating." "I would never talk to a machine," was an often heard comment 10 years ago. Most of us were uncomfortable leaving a recorded message on an answering machine without the familiar interaction with a human respondent (some of us still are). Yet today, telephone answering machines have become a way of life in urban upper-middleclass households. Lifestyle changes, resulting in many more households with no one home during the day, certainly contributed to their acceptance; but so did the familiarity from more than a decade of use. Now, many people who would not use an answering machine 10 years ago consider them necessities. As one friend put it, "I would much rather leave a message on a machine than have to call back again and again. In fact, I'm annoyed now when someone does not have an answering machine on the line." Objections that we hear today about cellular telephones and interactive voice response systems may similarly be transitory, so that these services may find widespread acceptance in the 1990s.

On the other hand, the promised benefits from new technologies and services are often overstated when they are first introduced. The promotional hyperbole accompanying a new development may then lead to a counterreaction gainsaying its advantages. The expansion of television program choices brought about by cable represents a good example. During the 1970s, cable proponents often talked about 100 or more new television channels, bringing subscribers everything from grand opera to local zoning hearings. Extravagant hopes for specialized, high-brow fare did not materialize, and some attempts to develop arts and cultural channels, such as the failed CBS Cable venture, suffered heavy losses. This has prompted some critics to decry cable's "failure" to provide "true diversity" of program choices. However, these criticisms should not overshadow the fact that cable does deliver many specialized channels of programming that are not found on broadcast television. They include (among others) channels dedicated to news, sports, weather, financial information, children's programming, rock music, country music, and live coverage of the U.S. Congress. While these offerings may not satisfy the critics who want non-commercial, high-culture programming, they have clearly increased the range of program choices offered to U.S. television viewers. Looking ahead, we should allow neither the promotors' overblown claims, nor the critics' cynicism, to obscure the real benefits that may come from the new products and services of the next decade.

Regulatory Influences on Innovation

The pace of innovation in telecommunications also depends strongly on regulatory policies. Despite the sharp turn toward deregulation during the past decade,[13] telecommunications remains highly regulated compared with other high-technology industries. Under the authority of the Communications Act of 1934, the Federal Communications Commission (FCC) grants cellular telephone licenses, approves telephone company (telco) investments in long distance transmission facilities,[14] allocates and assigns portions of the radio spectrum for commercial uses, sets detailed technical requirements for radio and television broadcasting, and generally develops, monitors and enforces technical standards for all spectrum uses.

Regulation generally is more effective in stable industries than in those where technologies and markets are rapidly changing. Under the Federal Administrative Procedure Act of 1946, the FCC must adhere to elaborate rulemaking procedures and give all interested parties time to review, comment on, and challenge proposed changes. These procedures significantly increase the time between development of an innovation and its commercial introduction, as well as the cost of bringing a new product or service to market. Cellular radio technology was developed by Bell Laboratories in the 1960s, and AT&T was ready to begin cellular telephone service by 1970,[15] but opposition by competing interests delayed its commercial introduction for some thirteen years. Developers of other communications services can tell similar stories.

Spectrum allocation and assignment probably raises the most formidable barriers to new service introduction. The FCC, charged with management of the broadcast spectrum, allocates specific frequency bands for specific uses, and then assigns particular frequencies within those allocations to particular users. This worked well in the 1930s when there were relatively few uses of the radio spectrum, but the explosion of technology and applications since World War II has created a hodgepodge of narrow, rigid frequency allocations. Various mobile communication services, for example, are allocated narrow bands of spectrum at different frequencies. With demand burgeoning, mobile communications suppliers are seeking to use bandwidth allocated for other services, such as FM radio subcarriers and portions of the UHF television spectrum. In this fragmented environment, a developer of a new digital paging service (e.g., sending stock quotes to portable, pocket-size receivers) may find it difficult to obtain compatible frequencies in different cities in order to distribute the service nationally. In hindsight, allocating a wider band of spectrum for mobile communications would help designers serve more users more efficiently.[15a]

In the interest of fairness to all prospective users, the FCC holds comparative hearings or lotteries to assign frequencies or grant licenses when there are many competing applicants. This process, however, may hamper entrepreneurial efforts to develop new communication services. For example, Airfone, Inc. started airplane to ground telephone service in 1977 under an "experimental" FCC license. The experiment proved successful, and a commercial demand for air to ground telephone service had clearly emerged by the early 1980s. The FCC, however, could not assign permanent use of the needed spectrum to the company that had pioneered the service. Instead, it proposed in 1984 to award the frequencies by lottery, allowing other companies to compete without incurring Airfone's development costs. Without the assurance of needed spectrum, Airfone was forced to sell out two years later to a larger company. Clearly, the FCC is faced with conflicting objectives of encouraging innovation and assuring equal opportunities for competing users of scarce spectrum. But the frequency assignment procedures now used tend to favor large, established companies over small, entrepreneurial developers of new services.

Technology's Impact on Policy

Technology influences communications policy in three principal ways: 1) it opens up new possibilities and expands the range of choices, which may require government action, approval or coordination; 2) it may bring unforeseen problems as it is widely adopted, demanding governmental response; and 3) it may also disturb the boundaries that legislators and regulators have drawn to separate communications services and the industries providing them.

Expansion of consumer choice comes about from improved variety, higher quality, and reduced cost: that is, more, better, and cheaper. Some technical advances offer wholly new choices not available before; as, for example, VCRs allowing viewers to record television programs for viewing at a more convenient time. Communications satellites introduced in the 1960s permitted trans-oceanic television transmission for the first time. Most technical developments, however, represent incremental improvements in quality, convenience and cost. Fiber optic systems improve the quality of telephone transmission; push-button telephones make calling more convenient (and open up possibilities for new voice-response services); telephone answering machines provide lower cost alternatives to answering services.

Here the policy goal is to maximize the flow of benefits from new technology while preventing market (or government) failures. For example, the FCC has decided to license two cellular telephone companies in each metropolitan area in order to insure the orderly development of cellular services while providing some benefits from competition. In other cases, such as high definition television, government policymakers may decide to play a leading or supporting role in setting standards. Defining the appropriate government role in standards is a tricky business, since the competitive marketplace may be better able to deal with system compatibility issues. Premature government action to set standards often proves counter-productive. Nevertheless, government action or coordination may be necessary to avoid policy gridlock which would inordinately delay the public benefits from technological development.[16]

Most technologies generate unforeseen problems when they are widely adopted, even if the initial uses seem relatively benign. Electronic bulletin boards, for example, make information widely available to PC users and encourage the exchange of messages among like-minded users. But bulletin boards have also created problems of fraud, libel, and obscenity.[17] Unscrupulous individuals have posted credit card numbers on electronic bulletin boards, thus encouraging fraudulent long distance calling or credit card purchases. In a few instances, electronic bulletin boards allegedly have been used to coordinate criminal activities by political fringe groups. In fact, the more effective is the new technology, the more likely it will be used for anti-social or illegal purposes. Public policy must then deal with these unexpected consequences of technological development and diffusion.

New technologies, however, probably influence communications policy most by creating conflicts between previously separated industries. The Communications Act of 1934 established clear boundaries between common carrier services, broadcast services, and other telecommunications services; and generally set up different rules and regulations for each sector. Rapid advances in communications technology during and after World War II began to play havoc with these regulatory boundaries. Microwave transmission made it feasible for companies to set up private communications networks in the 1950s, bypassing the telephone common carriers. Coaxial cable technology, in a continuingstate of development since the 1940s, permitted local cable television distribution in competition with the established broadcasters.

The development of electronic computing resulted in a collision between the information processing and telephone industries. The conflict was recognized in the 1956 AT&T antitrust settlement which prohibited AT&T and Western Electric from entering the burgeoning computer business. But technological development continued. Telephone switches essentially became computers, while computer manufacturers saw lucrative markets for their products in telecommunications. By 1969, when the FCC held its first Computer Inquiry, the technological convergence between the unregulated computer industry and the regulated telecommunications carriers was clear. This collision between two large technological industries, highlighted by the battles between AT&T and IBM, dominated much of communications policy-making during the 1970s and was a significant factor in the 1981 decision to break up the Bell System. Today, the convergence of telecommunications and computing is widely recognized, but its political and policy reverberations still echo.

Technology's march ensures similar collisions between entrenched business interests in the next decade. The development of high definition television threatens to leave over-the-air broadcasters behind cable television, direct broadcast satellite (DBS) and other technologies with fewer spectrum constraints. Fiber optic developments bring new competition between the cable and telephone industries for delivering video programming to the home. These developments are discussed in the next two sections.

II. HIGH DEFINITION TELEVISION

Television technology has advanced far beyond the National Television System Committee (NTSC) standard adopted in the U.S. nearly 50 years ago. High definition television (HDTV) promises picture quality comparable to 35mm film, with sound as good as that from a digital compact disc. The path from NTSC to HDTV in the United States, however, seems a long and bumpy one, with many obstacles along the way.

Japan's HDTV Developments

The Japan Broadcasting Corporation (NHK) has been the world leader in developing HDTV since the early 1970s. NHK has adopted a "studio production standard" with 1125 scanning lines and a 16:9 screen width-to-height aspect ratio. The NTSC standard is based on 525 scanning lines and a 4:3 aspect ratio. HDTV's wider screen, as much as its higher resolution, contributes to viewers' perception of a superior picture. Each HDTV frame has more than five times as many picture elements as an NTSC frame.

The NHK HDTV studio production standard has been endorsed by the Society of Motion Picture and Television Engineers (SMPTE) and is now being used for movie and television production in North America. It offers picture quality comparable to 35mm film with the advantages of videotape for easy editing and fast turnaround. Chasing Rainbows, a mini-series presented by the Canadian Broadcasting Corporation, and

Innocent Victims, a TV movie produced for CBS, both were shot using HDTV videotape and then converted to the NTSC standard for broadcasting. The NHK 1125-line format seems likely to be used widely for movie and TV production in the 1990s. Within a decade, extensive HDTV production could encourage the studios to replace film prints with satellite distribution of HDTV movies for theater projection.

NHK demonstrated its HDTV system last summer by transmitting the Olympic games from Seoul, Korea to various demonstration sites in Japan via satellite. HDTV broadcasting to Japanese home viewers will begin in 1990 when Japan launches its BS-3 direct broadcast satellite (DBS) service. The DBS service will use a bandwidth compression technique known as MUSE to reduce the transmission bandwidth from the 30MHz required for the studio production system to about 9MHz. Japanese manufacturers will also introduce HDTV television receivers, VCRs and videodisc players commercially in Japan by 1990; and they have announced plans to bring this HDTV consumer equipment to the United States in 1991 or 1992.

The NHK HDTV system is not compatible with NTSC (or with the PAL or SECAM 625-line television standards widely adopted in Europe and other parts of the world). Viewers will have to purchase new, expensive receivers to get HDTV; prices in the early 1990s are expected to fall in the $3,000-3,500 range. Moreover, the 10MHz MUSE system planned for DBS broadcasting in Japan cannot be carried over the 6MHz U.S. broadcast television channels assigned by the FCC.

Alternative Systems for the United States

Various alternative advanced televison systems that would be more compatible with NTSC have been proposed for the United States.[18] Most of these alternatives are in the conceptual or prototype stages and have not yet been field tested. "Advanced television" (ATV) is a more appropriate rubric for these systems than HDTV, since some of them involve improving the 525-line NTSC signal without providing more lines of resolution or wider aspect ratios. Table 2 shows the leading ATV concepts proposed as of this writing.[19] They are listed in three categories: 1) systems that can be carried on a single 6MHz channel; 2) systems that require a second channel to augment the basic NTSC signal; and 3) wideband systems that require transmission bandwidths greater than 6MHz. Note that these are all analog systems. Digital transmission of HDTV, or of television signals in general, awaits the development of broadband fiber optic networks to the home (see next section).

One ATV approach is to use today's technology to clean up the 6MHz NTSC signal as much as possible. Some of the NTSC viewing defects or "artifacts" stem from the way color information was added in the 1950s to the black and white "luminance" information, resulting in interference between the luminance signal and the color signal. Other artifacts arise from the interlaced scanning adopted as part of the NTSC standard which requires two scans ("fields") to build each picture. Signal processing techniques can remove or minimize these artifacts, giving a cleaner NTSC signal for broadcast or cable transmission over a conventional 6MHz channel. Television receivers with the appropriate decoders, either built-in or added as an adaptor unit, will show an improved picture. But standard NTSC receivers will still be able to display the standard NTSC picture with relatively little degradation.[20]

Faroudja Laboratories has developed its "Super NTSC" system using improved scanning, filtering and signal processing techniques. A proprietary encoder must be added to every NTSC camera to give the improved signal, which then is transmitted over conventional 6MHz channels. Special receivers with line doubling, frame storage and signal processing circuits can display a 1050 line picture. This equipment has been built and successfully demonstrated, but it is expensive; coders and decoders now cost several thousand dollars each. Mass production could bring down the cost significantly, however, if the Faroudja Super NTSC system were widely adoped.[21] Matsushita and Hitachi have also developed systems of this sort to improve picture quality while still retaining the NTSC transmission standard.

Other systems proposed by the David Sarnoff Laboratory (formerly part of RCA, now owned by SRI International), the Del Rey Group, and NHK use different technical approaches to transmit wide aspect "improved NTSC" signals over a 6MHz channel. For each system, standard receivers would display a standard, perhaps sightly degraded NTSC picture. New receivers with signal processing circuits would show noticeably improved pictures.

Better HDTV pictures can be transmitted over a 6MHz channel by sacrificing NTSC compatability. NHK has proposed a "Narrow MUSE" system with 1125 scanning lines and a 16:9 aspect ratio providing, according to NHK's computer simulations, pictures superior to all the "improved" or "enhanced" NTSC approaches. MIT has proposed another 6MHz "bandwidth efficient" system incompatible with existing NTSC receivers.

The dual channel ATV systems augment the NTSC signal with additional picture and sound information carried on a second channel. Suitably equipped receivers process both signals and recombine them to display a higher resolution image. Prototype dual channel systems have been demonstrated by William and Karen Glenn of the New York Institute of Technology, North American Philips, and the Sarnoff Laboratory. Another dual channel system has been proposed by NHK as part of its MUSE family.

Finally, wideband systems such as the NHK's original MUSE and HDMAC-60, proposed by North American Philips, offer high quality HDTV transmission at bandwidths greater than 6MHz. They, of course, require new television receivers and are incompatible with NTSC. At present, these wideband approaches seem most appropriate for satellite transmission and HDTV videocassette distribution.

There is really no way to compare the systems listed in Table 2 at the present time. Only the NHK MUSE system has been fully developed and field tested. The others, including the variations on MUSE announced by NHK this spring, are still in the early development stages. Technically, one can say that systems using more bandwidth will generally provide better pictures and/or be more robust in various transmission environments. At the same time, maintaining compatability with NTSC will generally lead to less efficiency in bandwith utilization and/or compromises in picture quality. Dual channel systems require the greatest bandwidth but may not necessarily deliver the highest quality pictures. So far, comparisons among ATV systems have been made principally by computer simulation. Direct comparisons require operating hardware systems, which are still at least months and millions of R&D dollars away.

Policy Issues and Stakeholders

HDTV development in the United States raises all three kinds of public policy issues noted in the introduction:

1. What should government policymakers do, if anything, to make HDTV available to American viewers as rapidly and inexpensively as possible?

2. How should policy deal with the problems raised by HDTV's introduction, or the unintended impacts of its widespread adoption? For example, should U.S.'viewers' investment in NTSC receivers be protected in some manner?

3. How will the existing stakeholders be affected? What role should public policy play in mediating among their competing interests?

This last question demands particular attention, since HDTV policy involves many major players with substantial stakes:

- Broadcasters have the most at risk. Their 6MHz channel assignments and detailed technical regulations constrain them far more than other distribution media. Their worst-case scenario has U.S. viewers responding eagerly to MUSE or another incompatible HDTV system in the 1990s. Broadcast stations might then see their audiences and revenues erode as cable and DBS distribute HDTV programs. The broadcast industry thus hopes to find some consensus favoring an NTSC compatible, 6MHz solution, or a dual channel approach. But how would broadcasters obtain a second channel for HDTV transmission? With today's technology, more broadcast channels could be squeezed into the UHF television band,[22] but the existing regulatory regime makes spectrum reallocation or reassignment excruciatingly difficult. Allocating more UHF bandwidth for television would be bitterly opposed by land mobile services and other prospective users of broadcast spectrum.

- Cable operators see HDTV as an opportunity to offer subscribers new services and further differentiate cable from broadcast programming. Although state-of-the-art technology can pack more than 70 channels on a single cable, most cable systems do not have unused capacity and would have to displace some existing programs to carry HDTV. Moreover, distortion, reflections and other transmission artifacts would show up much more strongly on HDTV than on NTSC signals. Cable operators thus will have to upgrade or technically "tune-up" their systems to carry HDTV. Still, cable companies have strong incentives to embrace HDTV when it appears in the U.S. market - both because HDTV may attract new subscribers and increase revenues, and because cablecompanies cannot afford to let videocassettes, DBS or the telephone companies overtake them as the distributors of HDTV to the home.

o Satellite system operators and programmers view HDTV as possibly the key to DBS success in the 1990s. Although DBS has not been successfully launched in the United States, with several prominent failures during the 1980s, it technically is an excellent way to distribute HDTV signals. DBS systems are not constrained by the 6MHz channels assigned to terrestrial broadcasters. Consequently, HDTV could enable DBS systems to make an end run around their broadcast and cable competitors. Home Box Office, RCA/GE, and Hughes Communications show interest in backing DBS systems which would include high definition programming. Japan, of course, will demonstrate the technical feasibility of transmitting HDTV via DBS beginning in 1990 or 1991.

- Telephone companies (telcos) see HDTV as a way to get into the video distribution business and justify building fiber optic networks to the home. Although currently prohibited from offering video services themselves (although not from building facilities and leasing capacity on them to others), the telcos do not want to be cut out of the video distribution business in the long term. They believe in a "single fiber solution" that would ultimately carry all voice, data, and video services to the home. Broadband fiber optic distribution systems could carry HDTV without the 6MHz bandwidth limitations of broadcast channels. Bell South demonstrated HDTV transmission over fiber optics at the 1988 Democratic National Convention in Atlanta, and several telcos hope to begin field trials of HDTV distribution as soon as the technology is commercially available (see following section for further discussion).

- U.S. consumer electronics manufacturers see HDTV as the vehicle to regain a substantial share of the TV equipment market lost to Japan over the last two decades. They strongly favor adopting a U.S. standard for HDTV different from the NHK standard, so that Japanese manufacturers will not have a head start in building HDTV receivers for the U.S.

market. "Let's not lose another market to Japan," say William and Karen Glenn in a recent New York Times article:

> "The American market for HDTV sets is estimated to be as big as $20 billion a year by 1997. The big question is where will those sets be made."[23]

European consumer electronics manufactureers such as Philips also favor rejecting the NHK HDTV standard so they have a better shot at the U.S. market.

- Television viewers are the stakeholders most directly affected, since they have some $50 billion invested in more than 150 million NTSC receivers. While the early adopters will benefit from better television pictures and sound, other viewers will not want a new HDTV system to obsolete existing program services or their investments in TV sets and VCRs.

No one in fact knows how eagerly U.S. television viewers will take to HDTV. The system developers in Japan, Europe and North America are optimistic that demand will swell once consumers compare HDTV pictures and sound with conventional television. The rapid market acceptance of compact disc audio players helps fuel that optimism. But there remain skeptics. A recent audience study by the Advanced Television Research Project at MIT reports:

> "Mass audience viewers prefer HDTV to NTSC in simultaneous, side by side comparisons. But the preference is not as strong as was expected and is highly influenced by the character of the programming and viewing conditions. We conclude that, to the eyes of the typical television viewer, the difference between NTSC and baseband HDTV is a subtle one highly dependent on environmental factors."[24]

The MIT study also concludes that "Viewers did not indicate a willingness to pay a substantial premium for an advanced video receiver based on what they saw in our tests." These results are not conclusive, as the MIT study makes clear, but they suggest that HDTV acceptance in the marketplace may not be as rapid as was the adoption of color television receivers and VCRs.

Setting HDTV Standards

Although HDTV has been under development since the early 1970s, U.S. stakeholders and policymakers did not seriously focus on HDTV issues until very recently. Today, most participants and observers agree that agreement on standards that provide for an evolutionary path from NTSC to HDTV is essential. And because the issues are so complex, with so many contending interests, the Federal Government, through the FCC, should take the lead in setting HDTV standards.

In November, 1987, the FCC created an Advisory Committee on Advanced Television Service. The Committee's Interim Report, released in June, 1988, recommends "establishing, at least ultimately, an HDTV standard for terrestrial broadcasting."[25] But the Committee faces formidable hurdles. Except for the Japanese NHK system, none of the advanced television systems listed in Table 2 has been tested thoroughly enough to warrant comparative judgments. Most are only technical concepts supported by computer simulations. Development and testing of even prototype hardware will not be feasible before 1989, and determining the full "spectrum impact" of the various alternatives may take at least another year.[26] By then, of course, the NHK system may be in commercial operation.

On September 1, 1988 the FCC issued a Tentative Decision that any broadcast HDTV service must be compatible with current NTSC receivers or simulcast with an NTSC signal on separate broadcast channels.[26a] This decision rejects MUSE or any over-the-air system that requires more than 6 MHz of continuous broadcast spectrum. However, HDTV transmission by cable, satellite, or other non-broadcast means could use a non-compatible system. The FCC's Tentative Decision also asks for comments on whether additional spectrum should be made available to broadcasters for augmented HDTV transmission on a second 3MHz or 6MHz channel. Comments are due by October 31, 1988.

In the short term, then, the FCC is the principal policy focus for developing agreement on HDTV standards and evolutionary paths. The broadcast industry is supporting an Advanced Television Test Center to test the various HDTV concepts under development. Broadcasters are also funding an Advanced Television Research Program at MIT. The cable industry has its own HDTV committee and has established a Cable Television Laboratory to pursue HDTV developments. These programs could bring about the first real cooperation between the broadcast and cable industries in seeking a consensus on ATV and HDTV.

In the longer run, policy should aim at making HDTV available to viewers without the constraints imposed by the 50-year-old NTSC system. The FCC seemed to favor this viewpoint in its Notice of Inquiry on advanced television systems:

> "To the extent that such an approach is both technically feasible and economically efficient, we now incline toward the view that, in the event we establish improved broadcast television systems, they should be implemented in a manner that allows eventually for the complete replacement of the NTSC, so that the benefits of improved off air television service may be enjoyed by the nation's viewers generally."[27]

This approach will likely require a two-stage evolution. The first stage would include one or more advanced television transmission systems compatible with existing NTSC receivers but offering enhanced picture and sound quality to those who purchase new equipment. Developing a "smart" television receiver may be the key the key to this scenario. The smart receiver would contain signal processors that could decode both NTSC and ATV signals. If the smart receiver has an "open architecture" that can process a variety of ATV signals, no single standard for ATV systems need be set in this first stage.[28] Competing single channel, dual channel and wideband systems could vie in the marketplace.

Over time, as the population of smart receivers increases, consumer preferences may decide among the competing ATV systems; or a high-efficiency HDTV standard could be agreed on that could be received on the smart receivers but not on NTSC sets. The transition to the NTSC incompatible HDTV system would represent the second stage. It would not likely occur for at least a decade, but it would herald the arrival of true high definition television for the 21st century.

III. BROADBAND DISTRIBUTION NETWORKS

The debate over high definition television may be only a preliminary skirmish before the main event: who will provide broadband services to businesses and homes as we approach the year 2000? The principal contenders are the telephone and cable television industries, with terrestrial and satellite broadcasters having significant stakes in the outcome as well.

Today, telcos and cable companies run separate businesses over separate networks. The telcos operate a switched, two way network that relies primarily on twisted pairs of copper wires to distribute voice and data services to home and business subscribers. Cable companies operate unswitched, one way video distribution networks based on coaxial cable. But these separate networks and businesses are converging as fiber optic distribution systems improve in performance and cost.[29] Telephone and cable companies well recognize that this new technology has the potential to carry all video, voice and data services on a single pair of glass fibers. Consequently, collisions between the two industries over who will own and operate fiber distribution networks seem inevitable in the 1990s.[30]

Telco Scenarios

Within telco networks, fiber optics are beginning to move into the "local loop" that runs from the telephone central office to the subscriber. Transmission between telco central offices -- the interexchange network -- is rapidly becoming digital and mostly fiber. Fiber optics are also used in "carrier systems" between telco central offices and distribution points nearer to subscribers. Bell South estimates that 50% of new subscriber installations use carrier systems and 50% of those carrier systems employ fiber optics.[31]

For now, the final links to the home -- the "last mile" -- are still copper wires. Wire pairs are less expensive than fibers and remain the cost efficient choice for voice and most data services. As shown in Table 3, home communications services such as security, utility metering, electronic messages, networking among personal computers, information retrieval, and slow-scan video do not require wide bandwidths or high data rates. These services can be provided over conventional telco copper wire local loops.32

But not full-motion video. Distributing standard NTSC television requires 6MHz of analog bandwith or 90 million bits per second in digital form. Processing of the video signal to reduce redundancies can cut the required bit rate by a factor of four or more; but even with data compression, television signals cannot be feasibly distributed over narrowband, copper wire distribution networks.

Telco scenarios thus include two ways to bring fiber optics to the home. The cost of fiber local loops may become competitive with those for conventional copper wire loops, or the demand for new video services may justify building fiber distribution systems.

On the cost side, Bell South has stated that by 1990, it can cost-justify installing fiber loops for new residential customers for voice telephone service alone[33]. Analyses by other telcos, consultants and university researchers do not support that assertion, however.[34] A 1987 study by Sirbu et al. at Carnegie Mellon University concluded:

> "it is clear that running fiber to the home, even assuming significant future reductions in component costs, is likely to remain more expensive than copper, where current loop plant costs are roughly $700 per subscriber."[35]

Moreover, optical fibers, unlike copper wires, cannot carry electrical current to make the telephone ring and provide other essential functions. Installing fiber optic local loops thus would require new telephones powered by batteries or the subscriber's electrical service. Long-term power outages could then result in loss of telephone service as well.

Fiber networks are more cost competitive if they can replace both the telephone wire and coaxial cable distribution systems, whose combined costs (in 1987 dollars) run about $1200-1500 per household. The cost of fiber distribution is much higher today, but costs are coming down steadily. By the mid to late 1990s, many telcos believe it will be as cheap to build one fiber distribution network as to construct two separate telco and cable systems.

On the demand side, it is not clear what new services beyond those listed in Table 3 will justify building fiber networks to the home. Coaxial cable systems are quite efficient for conventional television distribution, although the cable companies themselves are beginning to install fiber optics in their distribution networks (see below). Cable amplifiers now operate at bandwidths of 550MHz -- enough to carry 72 television channels per cable and more than current fiber technology can handle in either analog or digital form. Cable systems should be able to find room for high definition channels when a demand for HDTV emerges in the 1990s. Fiber systems are not required for HDTV distribution.

Nevertheless, a number of U.S. telcos have begun or plan field trials of direct fiber-to-home systems. They include BellSouth (Hunter's Creek and Heathrow, Florida), Southwestern Bell (Leawood, Kansas), Bell Atlantic (Perryopolis, Pennsylvania), U.S. West (St. Paul, Minnesota) and GTE (Cerritos, California). These trials generally use a mix of wire, cable and fiber to carry different services; or they use separate fibers for voice and video. Broadband digital switching is not yet commercially available that could integrate all voice, data, and video services on a single fiber to the home.[36]

Yet the telcos' ultimate goal is to provide broadband digital switching capacity so that their customers can connect to any source of video, voice or data.[37] This would enable them to offer "video-on-demand" services qualitatively different from those available on unswitched cable networks. With video-on-demand, subscribers could order up a movie or another video program at any time, rather than be tied to a programmer's schedule. Bellcore researchers have outlined the technical requirements and projected costs of video-on-demand; and they appear optimistic that home viewers would prefer it to fixed schedule services or videotape rentals.[38] However, there is still no indication of real consumer demand for this kind of service. Video-on-demand, like Picturephone and videotex, may be yet another example of technology searching for a market.[39]

The situation seems quite different for business customers. Many businesses already generate enough voice and data traffic to justify building digital fiber optic links on the basis of current fiber costs. Future business video services, such as videoconferencing or the exchange of video information among business workstations, may bring additional demand for broadband capacity. The regional Bell companies and other local telcos are moving quickly to provide broadband, mostly fiber links to their large business customers before they are bypassed by AT&T or one of the other long distance carriers. As a consequence, digital fiber optic distribution networks will reach businesses long before they get to the home.

Cable Scenarios

Cable companies, too, are testing fiber optics for video distribution.[40] For some years cable systems have used optical fibers for "transportation" links of several miles or more from satellite receivers to cable headends. Now fibers are proving competitive with coaxial cable and microwave for linking headends with the multiple hubs of large cable systems. These applications employ analog frequency modulation (fm) techniques to carry 12-24 television signals per fiber, so that handling the 40 to 80 channels typically required means running several fibers in parallel. Fiber systems are chosen not for their capacities (coaxial cables can carry more channels), but because they do not require amplifiers and thus deliver higher quality signals to the hubs.

Deploying optical fibers in the distribution plant from hub to subscriber (trunks, feeders, and drops) presents greater technical and cost challenges. Coaxial cables now carry up to 72 television signals in analog form with amplitude modulation (am) that is compatible with existing line amplifiers and home television receivers. Fiber systems operate best with digital or analog fm transmission requiring no amplifiers. Consequently, replacing cable with fiber implies converting the signals at some point from digital or analog fm to analog am, which is too expensive today to be competitive with conventional cable distribution.

R&D efforts to perfect multi-channel am transmission over fiber, and to reduce the cost of fm/am conversion, are underway in several laboratories. If successful, they would quickly lead to greater use of fiber for video distribution. ATC, the nation's second largest cable company which is 80% owned by Time Inc., has advanced a plan to use fiber optics instead of coaxial cable in trunk lines. The principal advantage is to reduce the number of amplifiers between the cable headend and the subscriber in order to improve reliability and deliver better quality television signals, rather than to increase the number of channels. The ATC design, like others under consideration by the cable industry, does not bring fiber optics all the way to the home. Coaxial cables are still used for the last links to the subscriber. It does, however, represent a significant step toward eventual all-fiber networks for video distribution.

Policy Issues

Technology -- in this case the development of broadband fiber optic networks -- is again breaking down the legal and regulatory barriers between industries. In the short term, the principal policy issues involve the telco/cable crossownership prohibitions. Current FCC rules forbid telephone companies from providing cable television service within their telephone service areas. This crossownership prohibition was made statutory in the Cable Communications Policy Act of 1984. In addition, the Modified Final Judgment (MFJ) in the consent decree governing the breakup of the Bell System forbids the seven Bell Regional Holding Companies from providing cable service anywhere.

Telephone companies can build video distribution networks within their service areas and then sell or lease them back to franchised cable operators. As one example, Ohio Bell (part of Ameritech) recently constructed a fiber optic network linking the headend and hubs of the cable system serving Cleveland, Ohio. Ohio Bell does not operate the cable system or provide any video or other services over these facilities.

Other telcos are seeking to buy cable systems outside their telephone service areas.[41] The telcos do not want to be permanently left out of the video distribution business, especially when they see fiber optic developments eventually making feasible a single, integrated network for all voice, data and video services. The several telco field trials mentioned above indicate their interest in fiber-to-home distribution. In a controversial decision bitterly opposed by the cable industry, the FCC recently granted GTE a waiver of its crossownership rules to build a coaxial cable and fiber distribution system in Cerritos, California. Another company (partly financed by GTE) will operate the cable system, but GTE will offer "experimental" services like video-on-demand over the fiber network.[42]

The telcos thus are seeking to overturn the current regulatory, legislative and judicial crossownership provisions. They contend that consumers are being denied innovative services such as video-on-demand, and that cable companies should not have a monopoly lock on local video distribution. In response, the cable industry points to competition from videocassettes, over-the-air broadcasting and new DBS systems. Telco oponents also argue that the crossownership rules arose from past telco abuses of their monopoly power. Telco dominance of local exchange services will continue into the 21st century.[43] With the telcos having significantly more resources than cable companies (in 1987, each of the seven Bell RHCs earned more profit than the entire U.S. cable industry), the potential for unfair competition still exists.

Yet the telco attacks on the crossownership prohibitions have shown substantial success in recent months. The FCC has proposed broadening waivers of its cable/telco crossownership rule to encourage telco competition with cable companies in video programming as well as distribution. The Commission also proposes recommending to Congress that it rescind the crossownership ban in the Cable Communications Policy Act of 1984.[44] The National Telecommunications and Information Administration (NTIA) has recommended that telcos be permitted to provide "video dial tone"; that is, to lease video channels on a common carrier basis to any program supplier, not just to franchised cable operators.[45] The video common carrier concept was first proposed by the White House Office of Telecommunications Policy in 1974.[46]

The crossownership issue will be high on the communications policy agenda in the years ahead. Technically, in the short run through the mid-1990s, video and voice/data networks will remain largely separate. Video programs will be distributed on coaxial cables or optical fibers with little, if any two-way switching capacity. Few economies of scale will exist in operating the video distribution network and switched wire network jointly. But the problem of possible cross subsidy persists.

Dealing with the telco cross subsidy issue should be far more tractable outside their telephone service areas than within. Consequently, removing the telco/cable crossownership prohibitions outside telco service areas seems an appropriate policy goal. Within their telephone franchises, however, the burden of proof should remain with the telcos to avoid cross subsidies or other unfair competitive practices before the crossownership rules are changed.

In the longer run, the development of switched fiber optic broadband distribution networks poses significant structural issues for U.S. telecommunications policy. If the technology and economics favor a single, integrated fiber network, who should own and operate it? Should it be run as a common carrier (the current telco model), or with programming and content rights protected under the First Amendment (the cable model)? If video, voice and data services are combined on a single network, who should regulate it? Can video distribution continue to operate under municipal franchises, while voice and data services remain under state PUC jurisdiction? Policymakers still have a few years to consider these questions, but they will not go away. Policy should encourage field trials that include the active participation of both telcos and cable operators; and that make their findings about technical performance, cost, and demand publicly available.

Yet even if an integrated fiber network shows economies of scale, the economies may not be sufficient to justify monopoly provision of local distribution services. The benefits from technical and cost economies of scale must be weighed against the benefits of price and service competition. Unless the case for a single, integrated distribution network is overwhelming, the nation may be better served by two (or more) competing communications systems to the home.

IV. TECHNOLOGY AND POLICY CHOICES

High definition television and broadband fiber optic distribution networks illustrate the complex ways in which technologies interact with communications policy. From many industry perspectives, HDTV technology is an unwanted visitor from Japan that is upsetting the relatively stable relationships established in the United States among television programmers, distributors, and equipment manufacturers. HDTV shows how communications policy issues get tangled up with issues surrounding trade, employment, and industrial policies. It also exemplifies how technology develops more rapidly than policy can respond. Policymakers usually find themselves behind the power curve, struggling to catch up. Yet in this example, policy can influence technological development if it can be developed in a timely way, before the technical systems are locked in. Agreement on technical standards and an evolutionary path toward HDTV in the next two or three years can accelerate the consumer benefits from the new technology.

The development of fiber optic distribution networks well illustrates how technology brings formerly non-competing industries into contention. Fiber optic systems probably will not replace copper wires and coaxial cables, any more than television has replaced radio or satellites have replaced microwave transmission. Older technologies faced with new competition often find new niches. But fiber optic developments challenge the different regulatory regimes that have been established at the federal, state, and local levels for telephone and cable television services.

Most of all, technological change makes life difficult for policymakers, as well as for the contending interests they must deal with. Nowhere does Schumpeter's "gale of creative destruction" blow more fiercely than in telecommunications. The technical options often seem overwhelming even to experts in the field, reminding them of the maxim proffered by a Chinese military commander two millenia ago:

> "In any situation where there are 36 alternative courses of action, running away is always the best."

Policymakers faced with technological decisions may feel like heeding the commander's advice. But running away from technology seems neither feasible nor desirable. Even though policymakers often must mediate among the interests that technology has thrown together, the problems need not have zero-sum solutions. Unlike the situation in many other policy fields, technological change in communications usually generates more gains than losses, with users generally among the winners.

Table 1 ADOPTION OF COMMUNICATIONS TECHNOLOGIES

Technology	Approximate Year of Commercial Introduction	1 Million U.S. Households by	20% of U.S. Households by	50% of U.S. Households by	No. of Years from Introduction to 50% of Households
Telephone	1876	1899	1907	1946	70
Pushbutton telephone	1964	n.a.	1976	1986	22
Tel. answering machine	1970	1982	1988	x	x
Cellular telephone	1983	1988	x	x	x
Television (black & white)	1946	1949	1950	1953	7
Color TV	1961	1963	1967	1972	11
VCR	1975	1979	1985	1988	13
Cable TV	1948	1964	1979	1988	40
Satellite TV receiver	1976	1985	x	x	x
Compact disc player	1983	1986	x	x	x
Personal computer	1976	1981	1986	x	x

n.a. - not available
x - not yet achieved

Sources: Historical Statistics of the United States; Electronic Industries Association; AT&T Bell Laboratories; Paul Kagan Associates

Table 2 ADVANCED TELEVISION TRANSMISSION SYSTEMS

System Name and Developer	Scanning lines per frame	Aspect Ratio	Transmission Bandwidth (MHz)	Compatible with NTSC receivers?	Status
Single Channel Systems					
Super NTSC Faroudja Labs	525/1050	4:3	6	yes	demonstrated
HD-NTSC Del Rey Group	525	5:3	6	yes*	computer simulation
ACTV I Sarnoff Lab (RCA/GE/SRI)	525	5:3	6	yes*	computer simulation
MUSE 6 Japan Broadcasting Corp. (NHK)	525	16:9 or 4:3	6	yes*	computer simulation
Narrow MUSE NHK	1125	16:9	6	no	computer simulation
Dual Channel Systems					
VISTA Glenn/NYIT	1125	16:9	6+3	yes*	prototype demonstrated
HDS-NA North American Philips	1050	16:9	6+3	yes*	prototype demonstrated
ACTV II Sarnoff Lab	1050	5:3	6+6	yes*	computer simulation
MUSE 9 NHK	1125	16:9 or 4:3	6+3	yes*	computer simulation
WideBand Systems					
MUSE NHK	1125	16:9	10	no	to be introduced commercially in Japan in 1990
HDMAC-60 North American Philips	1050	16:9	9.5	no	prototype demonstrated

*ATV signals on NTSC receivers will show some degradation from standard NTSC, and the wider aspect ratios must be accommodated by masking the top and bottom of the screen or losing the side panels of the ATV picture.

Table 3 DATA COMMUNICATIONS REQUIREMENTS FOR HOME SERVICES

Service	Uncompressed data rate (kbps)
Telephone	64
Alarm services	0.1
Utility meter reading	0.1 - 1
Energy management	0.1 - 1
PC networks	1.2 - 64
Electronic mail	1.2 - 64
Information retrieval	1.2 - 64
Slow scan video	1.2 - 64
Video teleconferencing	64 - 1500
NTSC television	90,000
High definition television	200,000 +

Notes

1 Topics that explicitly are not covered in this paper include: 1) technology or industrial policy issues (e.g., should R&D tax credits be extended?); 2) international trade policy issues (e.g., policies to stimulate U.S. exports of telecommunications equipment); and 3) military developments or military applications of communications technology, except as they affect civilian policies.

2 For a good discussion of both "hard" and "soft" technologies, and their network implications, see Richard Vickers and Toomas Vilmansen, "The Evolution of Telecommunications Technology," IEEE Communications Magazine 25, No. 7 (July 1987) 6-18.

3 I. E. Sutherland, C. A. Mead and T. E. Everhart, Basic Limitations in Microcircuit Technology (Santa Monica, CA: The Rand Corporation, R-1956-ARPA, November 1976).

4 Kenneth J. Macleish, "Mapping the Integration of Artificial Intelligence into Telecommunications," IEEE Journal on Selected Areas in Communications 6, No. 5 (June 1988) 892-898.

5 The technical literature on ISDN is already enormous and growing rapidly. For introductory technical discussions, see the "Special Issue on Integrated Services Digital Network," IEEE Communications Magazine 24. No. 3 (March 1966); and Manu Malek,

"Integrated Voice and Data Communications Overview," IEEE Communications Magazine 26, No. 6 (June 1988) 5-15.

6 Kenneth L. Phillips, "ISDN'S Built-In Problems," Telecommunications (October 1987) 55-58.

7 Steven E. Minzer, "Toward an International Broadband ISDN Standard," Telecommunications (October 1987) 94-106; Richard J. Solomon, "Broadband ISDN," International Networks (September 1987). As a first step toward the broadband ISDN concept, the U.S. telephone industry has adopted a Synchronous Optical Network (SONET) standard of 155.52 Mbps.

8 For example, a single CD-ROM can store the equivalent of a year of the Los Angeles Times.

9 H. C. Willard, "High Technology Improves Cable Economics" (Portsmouth, VA: General Electric Company, Consumer Electronics Business Operations, October 1984).

10 Peter W. Huber, The Geodesic Network: 1987 Report on Competition in the Telephone Industry, (Washington, D.C.: Antitrust Divison, U.S. Department of Justice, January 1987) 1.3.

11 Data on cable and VCR growth are from Cable TV Investor and The VCR Letter (Carmel, CA, Paul Kagan Associates, 1988)

12 CB radio seems a striking counter-example. It was adopted largely by middle and lower-middle class consumers in the early 1970s, but never achieved true mass market acceptance.

13 Michael J. Marcus, "Technical Deregulation: A Trend in U.S. Telecommunications Policy," IEEE Communications Magazine 25, No. 1 (January 1987) 66-68.

14 State regulatory authorities, of course, must also approve telco investments in their jurisdictions.

15 William C. Jakes, Jr., "New Techniques for Mobile Radio," Bell Laboratories Record (December 1970) 327-330.

15a See Kenneth Gordon and Alex D. Felker, A Framework for a Decentralized Radio Service, (Washington, D.C.: Federal Communication Commission, Office of Plans and Policy Working Paper, October 1983). Similar recommendations for the UK are contained in CSP, International, Deregulation of the Radio Spectrum in the UK (London: HMSO, March 1987).

16 Stanley M. Besen and Garth Saloner, "Compatibility Standards and the Market for Telecommunications Services,", (Santa Monica, CA: The Rand Corp., P-7393, February 1988); John Carey, "The Downside of Deregulation: A Case Study of U.S. Teletext," Electronic Publishing Revue 6, No. 1 (1986) 37-40.

17 Eric C. Jensen, "An Electronic Soapbox: Computer Bulletin Boards and the First Amendment," Federal Communication Law Journal 39, No. 3 (October 1987) 217-258.

18 For two excellent recent discussions of alternative HDTV systems, see "Special Issue on High Definition Television," IEEE Transactions on Broadcasting BC-33 (December, 1987); and "Special Issue on Advanced Television Systems," IEEE Transactions on Consumer Electronics 34 (February, 1988).

19 Other ATV systems have been proposed by MIT, AT&T Bell Laboratories, CBS, Hitachi, Matsushita, and Scientific Atlanta, among others. Eureka, a European consortium, is developing its own HDTV production system and transmission system that will not be compatible with any of systems listed in Table 2.

20 The degree to which any of these systems will degrade the picture shown on standard NTSC receivers is still very much in question.

21 The Faroudja system will be field tested on two California cable networks beginning this fall.

22 Local UHF channel assignments have been widely separated so that television receivers could tune to one UHF channel without interference from another. The channel spacing rules are known as the UHF "taboos." Technical improvements in TV tuners now

could permit eliminating many of these taboos. See Gary S. Kalagian, Robert P. Eckert and William A. Daniel, Advanced Technology UHF Receiver Study Part 2. Effect on UHF Television Allotments, (Washington, D.C.: Federal Communication Commission, FCC/OST R-84-1, March 1984.)

23 William E. Glenn and Karen G. Glenn, "Let's Not Lose Another Market to Japan," New York Times (March 6, 1988).

24 W. Russell Neuman, The Mass Audience Looks at HDTV: An Early Experiment, paper presented at the Annual Convention of the National Association of Broadcasters, Las Vegas, Nevada (April 11, 1988) 11.

25 Federal Communicaions Commission, Interim Report of the FCC Advisory Committee on Advanced Television Service. (Washington, D.C., June 16, 1988.)

26 Michael J. Marcus, personal communication.

26a Federal Communications Commission, Tentative Decision and Further Notice of Inquiry, MM Docket No. 87-268 (Washington, D.C.: September, 1988).

27 Federal Communications Commission, Notice of Inquiry MM Docket No. 87-268. (Washington, D.C.: August 20, 1987.)

28 Technically, an "open architecture" smart television receiver requires a standard interface between the signal decoders and the picture display. Different decoders might be needed to process different ATV signals distributed over different media, but the display (the most expensive part of the TV receiver) would be able to accept any decoded signal. See William F. Schneider, "HDTV Technology: Advanced Television Systems and Public Policy Options," Telecommunications (November 1987) 37-42; Indra Paul, Advanced Television System Model," Bell Communications Research (June 6, 1988) (Paper submitted to the Systems Subcommittee of the FCC Advisory Committee on Advanced Television Service.)

29 Several recent issues of the IEEE Journal on Selected Areas in Communications provide further technical background on fiber optic and broadband network development. See "Special Issue on Fiber Optics for Local Communications," SAC-3, No. 6 (November 1985); "Broadband Communications Systems," SAC-4, No. 4, (July 1986); "Special Issue on Fiber Optic Systems for Terrestrial Applications," SAC-4, No. 9 (December 1986); "Switching Systems for Broadband Networks," SAC-5, No. 8 (October 1987); "Fiber Optic Local and Metropolitan Area Networks," 6, No. 6 (July 1988); "Photonic Switching," 6, No. 7 (August 1988).

30 Walter S. Baer, "Telephone and Cable Companies: Rivals or Partners in Video Distribution?," Telecommunications Policy 8, No. 4 (December 1984), 271-289; Tom Valovic, "The Rewiring of

America: Scenarios for Local-Loop Distribution," Telecommunications (January 1988), 30-36.

31 M. Gawdun, "Lightwave Systems in the Subscriber Loop," Telecommunications (May 1987) 65.

32 Thus telco implementation of ISDN services to both homes and businesses has assumed copper wire local loops. Of course, fiber optic networks with data rates greater than those shown in Table 3 could offer faster response times for information retrieval, facsimile and other services; but whether consumer demand would justify the higher cost remains questionable.

33 R. K. Snelling and J. Chernak, "The Revolution in the Loop Network: The Evolving Broadband Technology, Proceedings ISSLS 1986 (September 1986) 210-215; R. K. Snelling and K. W. Kaplan, "All Fiber Networks - A Reality by 1988-89," Conference Record, ICC 1987 (June 1987) 44.1-44.1.5; Scott Ticer, "Jeno Paulucci's Dream: Bring Fiber Optics Home," Business Week (September 21, 1987), 34-35.

34 George P. Lynch, "Illinois Bell Studies Local Loop Fiber," Telephony (August 10, 1987) 62; Howard Rausch, "Telephone System in the Year 2000," Lightwave (August 1988) 17-18. Future cost comparisons are difficult because they depend so crucially on such assumptions as future copper and fiber costs, maintenance

costs and household densities. For a good illustration of how sensitive the results are to these assumptions, see Myron Keller, "Economic Considerations of Fiber in the Loop Plant," (St. Louis, MO: Southwestern Bell, February 1988).

35 Marvin Sirbu, Frank Ferrante, and David Reed, "An Engineering and Policy Analysis of Fiber Introduction into the Residential Subscriber Loop," (Pittsburgh, PA: Dept. of Engineering and Public Policy, Carnegie Mellon University, May 1988), 21-22.

36 Progress in fast packet switching techniques looks promising, but broadband digital switches that are cost competitive for residential services may be a decade away. See "Switching Systems for Broadband Networks," op.cit., especially D. R. Spears, "Broadband ISDN Switching Capabilities from a Services Perspective," 1222-1230; and P. E. White, J. Y. Hui, M. Decima, and R. Yatsubashi, "Guest Editorial - Switching for Broadband Communications," 1217-1221; and S. Cheng and L. Wu, "Reconfigure in a Flash," Telephony (May 18, 1987) 76.

37 An interim step could be a hybrid system delivering digital voice and data, along with analog video. See M. F. Mesiya, "Implementation of a Broadband Integrated Services Hybrid Network," IEEE Communications Magazine (January 1988) 34-43; Fred Dawson, "Three RBOCs Set Local Loop Tests of New Technology," Lightwave (August 1988) 1,6.

38 C. N. Judice, E. J. Addeo, M. I. Eiger, and H. L. Lemberg, "Video on Demand: A Wideband Service or Myth?" *Conference Record ICC 86* (June 1986) 1735-1739.

39 A. Michael Noll, "Videotex: Anatomy of a Failure," *Information and Management* 9 (1985) 99-109.

40 Fred Dawson, "Seeing Fiber Optics in a New Light: Cable's Mindset Beginning to Change," *Cablevision* (October 12, 1987) 48.

41 "Cable Wrestles With Possible Pactel Entry Into Business," *Broadcasting* (July 4, 1988) 32-34.

42 "GTE Details Fiber Plans for Cerritos," *Broadcasting* (June 27, 1988) 66-68.

43 Large business customers have competitive alternatives to local exchange carrier (LEC) services, but "[r]esidential and small business users still have few practical alternatives to LEC lines for short-haul transmission." Peter Huber, *The Geodesic Network*, op.cit. 2.23. Cable systems, in particular, do not have the two-way switched facilities in place to compete with the telcos for residential and small business telephone service.

44 "FCC Wants to Loosen Cable-Telco Prohibitions," *Broadcasting* (July 25, 1988) 31-32.

45 U.S. Department of Commerce, National Telecommunication and Information Administration, Video Program Distribution and Cable Television: Current Policy Issues and Recommendations (Washington, D.C.: NTIA Report 88-233, June 1988).

46 Clay T. Whitehead, Chairman, Report to the President (Washington, D.C.: The Cabinet Committee on Cable Communications, Office of Telecommunications Policy, 1974). See also Ithiel de Sola Pool, Technologies of Freedom (Cambridge, Massachusetts: Harvard University Press, 1983, 151-188.

REGULATORY INCENTIVE MECHANISMS AND POLITICAL ACTION[1]

David P. Baron
Stanford University

I. Introduction

Recent developments in economic theory have generated a considerable literature on the design of regulatory incentive mechanisms that address the informational and monitoring problems that complicate economic regulation.[2] This research has paralleled innovation in regulatory practice such as the incentive mechanism used to regulate the Public Service Company of New Mexico, the "RPI - X" system used to regulate British Telecom, and the price cap mechanism proposed by the FCC for regulation of AT&T. The New Mexico experiment succumbed to political pressure,[3] and the FCC plan has met a storm of political opposition in the Congress. The British system is still in effect, although political pressure from consumers has resulted in the creation of a separate cap for residential services.[4]

The fact that the price cap system has been in effect for nearly five years in the U.K. may in part be due to the parliamentary form of government which aligns the executive and the control of the Parliament thus to some extent insulating innovative policies from direct political pressure. In the U.S., the executive and the Congress are not necessarily aligned, and even when the president and the congressional majority are of the same party, members of the Congress often act with a degree of independence that reflects the interests of their constituents rather than a common policy. For example, the FCC was unable to implement its original access charge plan because of congressional pressure from both sides of the aisle. Without enacting legislation, the Congress was successful in delaying the implementation of the plan and in reducing the monthly charges for residential services.[5]

---

[1] This research has been supported by National Science Foundation Grant No. SES-8808211.

[2] Baron (1987), Besanko and Sappington (1987), Caillaud, Guesnerie, Rey and Tirole (1988), and Sappington and Stiglitz (1986) have surveyed this literature with an emphasis on Bayesian mechanisms. Non-Bayesian mechanisms are presented in Sibley (1988) and Vogelsang (1988).

[3] See Helman (1984).

[4] See Bhattacharyya and Laughhunn (1987) and OFTEL (1988).

[5] See Ferejohn and Shipan (1988) for an analysis of the political action associated with recent innovations in telecommunications regulation. Noll (1987) and Romer and Rosenthal (1986) have surveyed the literature on regulation and politics.

These regulatory innovations affect both distribution and efficiency, but it is the distributive consequences that generate the political action that limits the ability of regulators to implement efficiency-enhancing innovations.[6] Political action thus calls into question the implementability of regulatory incentive mechanisms. In particular, ignoring the political reaction to regulatory innovation may well limit the willingness of policy makers to consider innovations intended to improve regulatory performance. That is, if the designers of incentive mechanisms ignore the political action that may arise in response to the implementation of such a mechanism, regulators may well ignore the designer's advice. Conversely, by explicitly incorporating the likely political response to alternative regulatory innovations, regulatory theory may be more constructive. The purpose of this paper is to initiate the process of explicitly taking the political repercussions of regulatory innovation into account in the design of regulatory incentive mechanisms. This will be done by endogenizing the degree to which the regulator can commit to a long-term regulatory policy in the face of possible political action by interest groups.[7]

The incentive problem considered here is due to an asymmetry of information about the costs of the firm. Asymmetric information allows a firm to earn profits, or rents, if its costs are actually low, and the regulator seeks to control those profits as well as provide incentives for effort to reduce costs. An interest group is assumed to be able to observe the profit the firm earns in one period and to further its own interests by taking political action in the next period directed to forcing a revision in the regulatory policy. If its political action is successful in leading to revision in the regulatory policy, incentive regulation devolves into revenue-requirements or rate-of-return regulation. Information is thus the source of both the regulatory incentive problem and the political action problem.

Elsewhere I have argued that price cap regulation in the U.S. might be expected to devolve into rate-of-return regulation through revision of the price cap in response to political pressure resulting from the level of profits earned by the regulated firm.[8] This paper provides a more formal analysis of this issue. More specifically, the focus is on the ability of the regulator to make the credible long-term commitments to policy that are necessary to impart the proper incentives to the regulated firm. When credible commitment is not possible, the efficiency of regulatory incentive mechanisms is reduced as Laffont and Tirole (1987)(1988) have demonstrated. Baron and Besanko (1987b) argue that regulated industries are limited in the extent to which they are able to withdraw assets from regulated service and thus their response to the regulator's inability to commit to a regulatory policy is restricted.[9] Even in this case, however, the inability to commit results in inefficiency. The model analyzed by Baron and Besanko assumes that commitment is not possible but does not endogenize the forces that may make it impossible. The approach taken here is to endogenize the political action that may force revision in regulatory policies thereby making uncertain the credibility of commitments made by regulators. The regulatory mechanism considered here is different from price cap regulation in that only one service is considered and thus regulation involves the direct control of the price for that service.

---

[6] Stigler (1971), Posner (1974), and Peltzman (1976) address the distributive consequences of regulatory policies.

[7] The model studied here is in the class of Bayesian mechanisms, and most regulatory models assume that credible commitments can be made. Commitment is also assumed in non-Bayesian models, but it is not clear how that framework can be used to characterize behavior when commitment is not credible. For example, in non-Bayesian models information is revealed through performance of the firm, and political action to reduce the rents of the firm can thus be expected. What behavior can be anticipated and how a regulator should take that into account in the design of the mechanism is difficult to characterize in a non-Bayesian framework.

Political action may be initiated by interest groups or by political entrepreneurs both inside and outside political institutions. In modeling political action, it is desirable to represent both the interest groups and the institutions that resolve political issues. The purpose of the model, however, is not to be definitive about the nature of political action but instead is to represent that political action in a simple manner so as to facilitate investigation of the effect of that action on the design of a regulatory incentive mechanism. The model thus focuses on interest group competition without considering the richness of actual political institutions or the behavior of those occupying positions in those institutions.

With the advent of the deregulation movement, at least at the federal level the Congress rather than the regulators has been an impediment to the implementation of regulatory innovation intended to improve economic efficiency. Consequently, the focus here will be on political action that operates through the Congress with the regulator modeled as seeking to serve efficiency objectives.[10] Political action can be reflected either as political entrepreneurship by members of Congress or as political action by interest groups that puts pressure on members of Congress. The latter form seems to be the more pervasive and will be modeled here. The two interest groups considered are consumers and producers. Based on what the interest groups observe from regulatory performance, they take political action intended either to revise the regulatory mechanism or to preserve it. When revision takes place, the regulatory mechanism is altered to serve the interests of the "winning" interest group.

---

[8] See Baron (1988a) and Vickers and Yarrow (1988).

[9] In a model in which incentive problems arise because of sunk costs, Salant and Woroch (1988) characterize an equilibrium when neither the firm nor the regulator can commit credibly.

The regulator will be assumed to be an ideal bureaucrat interested only in fulfilling its mandate and to have the administrative authority to make credible commitments to long-term regulatory policies. Commitment does not mean that the policy does not change over time. Instead, it means that the regulator chooses a policy that will remain in effect for a specified time period. That policy may contain rules that change prices over time in a prescribed manner in response to known changes or in response to information revealed through performance. The regulator is assumed to have the power to commit to such a long-term policy as long as the political action of the interest groups acting through the Congress does not force a revision. In the absence of political action, commitment is achieved through constancy of commission membership, commission appointments made by the same president or governor, and administrative law requirements that make existing policies difficult to change administratively. Even though the regulator is able to commit to a long-term policy, political action by interest groups may be successful in generating congressional pressure or legislation that forces a revision in that policy.

In the model, political action results from interest group action based on information obtained from the observation of performance. In particular, the consumer interest group will be assumed to observe the profit or rents of the firm and to take political action to reduce those profits. If the political action is successful, the regulated price is reduced to a level that eliminates the rents given the information available to consumers. The industry has an opportunity to take political action to oppose the consumer interest group. Recognizing that this political competition may take place, the regulator must design a regulatory mechanism that takes into account the political action generated by that mechanism. The process represented by the model is illustrated in Figure 1.

---

[10] In practice, of course, regulators may seek to serve the interests of groups that will support them in their subsequent political or private careers.

The questions to be addressed in the paper are as follows. First, how does the anticipation of political action affect the prices set by the regulator? For example, does the regulator alter the prices set for the regulated services with the intent of diminishing the subsequent political action and hence the likelihood of revision of that policy. Second, does the regulator choose policies that limit the information about the firm's cost that the interest group can infer from its observation of regulatory performance? For example, the regulator may choose a policy that limits the information the consumer interest group has on which to base its political action. Third, what are the *ex ante* and *ex post* distributive and efficiency consequences of the anticipation of political action?

The answers provided to these questions are as follows. The regulator does not revise the price set in the first period, but in anticipation of the political action that will result in the second period, the regulator chooses a lower price than it would in the absence of political action. The regulator thus bows to anticipated political pressure by giving the interest group less incentive to take political action. Some political action is taken in equilibrium, however, and if it is successful, the regulatory policy devolves into revenue-requirements or rate-of-return regulation.

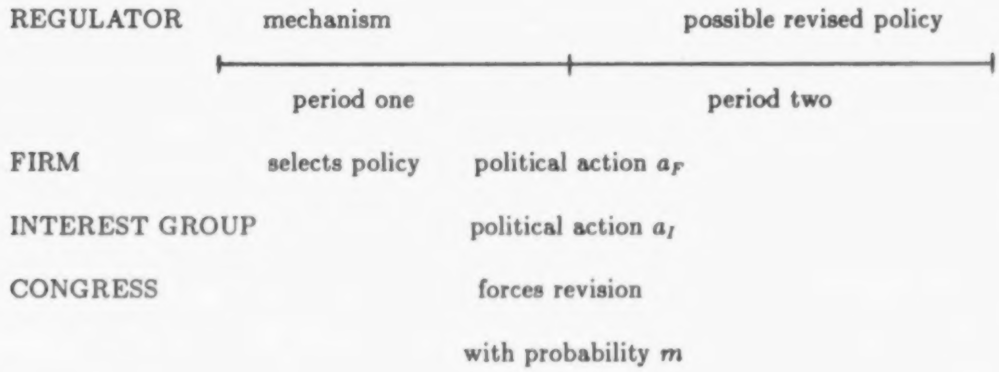
With respect to the distributive consequences, political action by the consumer interest group will be shown to reduce *ex ante* consumer well-being and to increase the profit of the firm. Similarly, political action by the firm to counter the consumer interest group will be shown to increase the *ex ante* well-being of consumers and to reduce *ex ante* profit. These counterintuitive results obtain because neither the consumer interest group nor the firm is able to commit not to take political action. That is, political rights are inalienable, so consumers are not able to prevent themselves from taking political action to improve their period-two welfare even though they understand that the anticipation of this political action will reduce their *ex ante* welfare. From a two-period perspective, the firm prefers that the consumer interest group take political action, but when the consumer interest group initiates political action in period two, the firm's *ex post* or period-two interests are to take political action to oppose the consumer group.[11] The inability to commit not to

---

[11] The same phenomenon will result in any finite-period model.

Figure 1

SEQUENCE OF ACTIONS


REGULATOR
mechanism
possible revised policy
period one
period two
FIRM
selects policy
political action a_F
INTEREST GROUP
political action a_I
CONGRESS
forces revision
with probability m

act in the future thus causes the *ex post* incentives of both consumers and the firm to be the opposite of their *ex ante* interests.

Even though consumers prefer not to take political action, a political activist may have a personal incentive to initiate action and to attempt to organize a political campaign to exploit the *ex post* interests of consumers. Similarly, a member of Congress may attempt to act as a political entrepreneur and claim credit for benefitting consumers *ex post* even if that harmed them *ex ante*; i.e., in the past. The rhetoric used to appeal to consumers would clearly be a claim that the rents of the firm were unearned and thus unwarranted. Such rhetoric is commonplace and may be effective among uninformed consumers. Furthermore, the media may find the issue attractive which may help reduce the cost of political entrepreneurship.

The next section presents the regulatory model and summarizes the properties of the optimal regulatory mechanism for the case in which commitment to a long-term policy is credible. The model of political action is introduced in Section III, and the two models are combined in Section IV. Extensions and alternative formulations are considered in Section V, and conclusions are offered in the final section.

II. A Regulatory Incentive Mechanism with Credible Commitment

A two-period model is sufficient to identify the effect of anticipated political action on the design of the regulatory incentive mechanism. The model incorporates a single regulator and a single firm. The two complications in the regulatory context are that the firm is better informed than is the regulator about the costs of providing the regulated service and that the actual decisions of the firm are not observable by the regulator.[12] Regulation takes the form of the control of prices for the regulated service. In the spirit of price cap regulation, the regulator is assumed not to use *ex post* observables, such as realized profit and costs, in the regulation.[13] The regulator thus only specifies prices for each period. Ideally, the regulator would like to base those prices on the costs of the firm, but not knowing those costs the regulator can rely only on indirect means of doing so. The regulator can accomplish this by designing a menu of price policies and having the firm selects one of those policies. The firm will make its selection based on its costs, and thus the pricing policy that will be implemented will depend on costs.[14] The task of the regulator is thus to design a menu of pricing policies, or a mechanism, that will induce the firm to choose a pricing policy for the costs that the firm actually has.

---

[12] The extension to the case of several regulated services is straightforward, using the approach in Sappington (1983), if the costs of providing those services are a function of a single unknown parameter.

[13] Alternatively, the regulator can be assumed to be unable to observe the performance of the firm. Laffont and Tirole (1986) and Baron and Besanko (1988) characterize optimal mechanisms in models in which the regulator can observe an *ex post* measure of performance.

The model is managerial in the sense that the management of the firm is represented as having preferences that diverge from those of its owners.[15] This is modeled as a risk-neutral manager making an effort decision with effort e_i reducing costs but carrying a disutility $D(e_i)$ for management, where $D(e_i)$ is strictly increasing and strictly convex. The effort and its disutility are assumed to be unobservable to both the regulator and the interest group. At the end of a period, however, the profit of the firm can be observed, so the interest group is able to obtain information about the firm's costs from the observation of its profit. The interest group is assumed to observe the mechanism and to draw inferences about costs from the observation of profits. That information can then be used as a basis for the interest group's political action.

Even though profit is observable at the end of a period, prices must be set at the beginning of the period, and thus they cannot be based on any observable. The observation of profits could be used by the regulator to set prices in subsequent periods, but as will be indicated, that information is redundant because the regulator can infer it from the firm's selection of a regulatory policy from the menu.[16]

In the formulation of the regulatory problem considered here, the private information of the firm allows the firm to earn rents relative to what it would earn if information were complete and symmetric. Those rents are a decreasing function of the price set for the output of the firm, and since consumer surplus is a decreasing function of the price, the regulator faces a trade-off between the control of those rents and consumer surplus. Since the regulator does not know the private information of the firm, its objective is necessarily specified as *ex ante*.

---

[14] The firm is assumed to be unable to withdraw assets from the regulated service as long as it is able to earn its opportunity profit. Baron and Besanko (1987b) consider multiperiod regulation under the assumption that the firm is required to participate as long as its profits are nonnegative given the information it reports to the regulator. In practice, a regulated firm is likely to have sufficiently large sunk assets that its continued participation is guaranteed. Laffont and Tirole (1987)(1988) consider the case in which the firm can quit the regulatory relationship.

[15] The managerial labor market, however, is not modeled.

[16] If the regulator could impose unlimited penalties based on the selected regulatory policy and the observed profit, the regulatory policy could be improved by basing it on performance.

The firm's marginal cost $C_i(\theta_i, e_i)$ in periods $i = 1, 2$ is assumed to be a function of a parameter θ_i that is known to the firm but unknown to the regulator and to the consumer interest group and of the effort of management. The marginal cost is specified as

$$C_i(\theta_i, e_i) = \theta_i - e_i, \quad i = 1, 2. \tag{1}$$

The parameters θ_i are related by a transition function θ_2 given by

$$\theta_2 = \theta_2(\theta_1, e_1). \tag{2}$$

To simplify the analysis, the focus will be on the case in which the parameters $\theta_1 = \theta_2 = \theta$ are the same in both periods. The information structure of the model thus is that the firm knows the parameter θ and the regulator and the consumer interest group only know the distribution function $F(\theta)$ and the corresponding density function $f(\theta)$. The density function is assumed to be positive on its support $[\theta^-, \theta^+]$.

Regulation takes the form of the control of prices, and two-part price structure is considered with a unit price p_i and a fixed fee t_i. The demand function $Q(p_i)$ is common knowledge and is assumed to depend only on the unit price, so the total fixed fee T_i is the product of the number of consumers and t_i. The regulator thus chooses a price policy (p_i, T_i) for each period $i = 1, 2$. The pricing policy and the profit earned by the firm are the only observables in the model. The profit, however, is observable only at the end of the period, and since prices must be announced to consumers at the beginning of the period, the observation of profit can only be used for inference about θ.

Under complete information, the optimal or first-best regulatory policy is to set the prices p_i^* equal to marginal cost or $p_i^* = \theta - e_i^*$, where e_i^* is the first-best effort which satisfies

$$Q(p_i^*) - D'(e_i^*) = 0. \tag{3}$$

Since the manager is risk-neutral, he will choose the first-best effort when the price is set at p_i^*.

The regulator is assumed to have the ability at the beginning of the first period to commit credibly to a policy for both periods. A regulatory pricing mechanism M is thus a collection $M = \{(p_i(\theta), T_i(\theta)), i = 1, 2, \forall\theta \in \Theta\}$.[17] In keeping with the openness of the regulatory process in the U.S., the regulatory mechanism is assumed to be public information.

The profit $\pi_i(\hat{\theta}, e_i; \theta)$ of the firm in period i is a function of its parameter θ, the effort e_i the manager chooses, and the pricing policy the manager selects from the menu of pricing policies it offers, where the choice variable $\hat{\theta} \in \Theta$ selects a policy $(p_i(\hat{\theta}), T_i(\hat{\theta})), i = 1, 2$.[18] The profit in period i is

$$\pi_i(\hat{\theta}, e_i; \theta) = p_i(\hat{\theta})Q(p_i(\hat{\theta})) + T_i(\hat{\theta}) - (\theta - e_i)Q(p_i(\hat{\theta})) - K, \tag{4}$$

where and the two-period profit Π is

$$\Pi = \pi_1 + \beta\pi_2, \tag{5}$$

where β is the discount factor. The profit in each period is observable, but the return R_i to the manager, which differs from profit because of the disutility $D(e_i)$ of effort, is unobservable. That return is given by[19]

$$R_i(\hat{\theta}, e_i; \theta) = \pi_i(\hat{\theta}, e_i; \theta) - D(e_i). \tag{6}$$

The two-period return $R(\hat{\theta}, e_1, e_2; \theta)$ of the manager is then[20]

$$R(\hat{\theta}, e_1, e_2; \theta) = R_1 + \beta R_2 = \Pi - D(e_1) - \beta D(e_2). \tag{7}$$

The revelation principle applies in this model, and thus there exists an optimal mechanism with the property that the firm chooses a pricing policy designed for its type or

---

[17] The following derivation of the optimal regulatory policy is similar to that presented in Baron and Besanko (1984).

[18] When the regulator can credibly commit and the cost parameter is the same in both periods, a single selection is sufficient for optimal regulation.

[19] The optimal managerial labor contract will make the manager the residual claimant of the profit of the firm. This formulation can also be obtained if the manager receives a fixed fraction ξ of the profit and $D(e_i)$ is replaced by $D^o(e_i) \equiv \xi D(e_i)$.

[20] The fixed cost K includes an *ex ante* wage for the manager. In a competitive market for the position of manager, the wage would be a function of the anticipated equilibrium political action but that connection will not be considered here.

$\hat{\theta} = \theta$.[21] Consequently, the regulator chooses a mechanism such that

$$(\theta, e_1(\theta), e_2(\theta)) \in \underset{(\hat{\theta}, e_1, e_2)}{\arg\max}\, R(\hat{\theta}, e_1, e_2; \theta). \tag{8}$$

The first-order condition for the firm's choice of effort is

$$Q(p_i(\theta)) - D'(e_i(\theta)) = 0, \quad i = 1, 2. \tag{9}$$

The revelation principle implies that, when the manager selects the pricing policy designed for the firm's type, his return $R(\theta) \equiv R(\theta, e_1(\theta), e_2(\theta); \theta)$, viewed as a function of θ, must have the property that

$$\frac{dR(\theta)}{d\theta} = -\big(Q(p_1(\theta)) + \beta Q(p_2(\theta))\big). \tag{10}$$

The return is thus a decreasing function of its type, so a manager of a higher cost firm earns lower rents on his information. The return of the manager can be expressed as

$$R(\theta) = \int_{\theta}^{\theta^+} \big(Q(p_1(\theta^o)) + \beta Q(p_2(\theta^o))\big) d\theta^o + R(\theta^+), \tag{11}$$

where $R_i(\theta) = \int_{\theta}^{\theta^+} Q(p_i(\theta^o)) d\theta^o$.[22] The return represents rents to the manager's informational advantage. Those rents result because the regulator must make incentive payments to induce the manager to select the pricing policy intended for the true cost parameter θ. There is no reason to give the manager of the highest cost firm any rents, so $R(\theta^+)$ can be set to zero.

The manager will participate only if the return $R(\theta)$ is at least as great as his opportunity cost, which will be taken to be zero. The participation condition thus is

$$R(\theta) \geq 0. \tag{12}$$

If $R(\theta^+)$ in (11) is set to zero, any incentive compatible mechanism satisfies (11). Regulation is also assumed to be characterized by the requirement that the return is nonnegative for each period, so $R_i(\theta) \geq 0$, $i = 1, 2$.

---

[21] The usual argument may be used to demonstrate that the revelation principle holds in the full model that includes the political action.

[22] The profit $\pi_i(\theta) \equiv \pi_i(\theta, e(\theta); \theta) = R_i(\theta) + D(e_i(\theta))$.

The regulator's objective is specified as the maximization of the *ex ante*, weighted surplus of consumers and the firm. Let $V(q_i), i = 1,2$, denote consumers' valuation of a quantity q_i, so consumer surplus S_i in period i is

$$S_i = V(q_i) - p_i q_i - T_i. \tag{13}$$

The weighted objective is $w = S_1 + \alpha R_1 + \beta(S_2 + \alpha R_2)$, where $\alpha \in [0,1]$ is the weight given to the return of the managers in the regulator's objective.[23] The *ex ante* objective W of the regulator thus is[24] [25]

$$W = \int_{\theta^-}^{\theta^+} [S_1 + \alpha R_1 + \beta(S_2 + \alpha R_2)] f(\theta) d\theta. \tag{14}$$

Substituting for the consumer surplus in (13) and the return R in (6) for $\hat{\theta} = \theta$ yields

$$\begin{aligned} W = \int_{\theta^-}^{\theta^+} \Big[& V(Q(p_1(\theta))) - (1-\alpha)\big(p_1(\theta)Q(p_1(\theta)) - T_1(\theta)\big) \\ & - \alpha\big[(\theta - e_1(\theta))Q(p_1(\theta)) + K + D(e_1(\theta))\big] \\ & + \beta\Big(V(Q(p_2(\theta))) - (1-\alpha)\big(p_2(\theta)Q(p_2(\theta)) - T_2(\theta)\big) \\ & - \alpha\big[(\theta - e_2(\theta))Q(p_2(\theta)) + K + D(e_2(\theta))\big]\Big)\Big] f(\theta) d\theta. \end{aligned} \tag{15}$$

Using the expression for the manager's return $R(\theta)$ to eliminate the revenue terms $(p_i Q(p_i) + T_i), i = 1,2$, yields

$$\begin{aligned} W = \int_{\theta^-}^{\theta^+} \Big[& V(Q(p_1(\theta))) - [(\theta - e_1(\theta))Q(p_1(\theta)) + K + D(e_1(\theta))] - (1-\alpha)R(\theta) \\ & + \beta\Big(V(Q(p_2(\theta))) - [(\theta - e_2(\theta))Q(p_2(\theta)) + K + D(e_2(\theta))]\Big)\Big] f(\theta) d\theta. \end{aligned} \tag{16}$$

From (16) it is evident that the objective function of the regulator is reduced by that portion $(1-\alpha)$ of the return to the manager that is not included in W.

---

[23] Baron (1988b) provides a political theory of the choice of that weight in a hierarchial structure in which a majority-rule legislature chooses α and the regulator then regulates a firm with private information about its costs.

[24] The discount factors for the firm and the regulator are assumed to be the same.

[25] The use of the return R_i in (14) rather than profit is equivalent to the *ex ante* wage of the manager covering his disutility of effort.

The expression in (11) for the manager's return can be used to eliminate $R(\theta)$ from (16). Taking the expectation of (11) and integrating by parts yields

$$\int_{\theta^-}^{\theta^+} R(\theta)f(\theta)d\theta = \int_{\theta^-}^{\theta^+} [Q(p_1(\theta)) + \beta Q(p_2(\theta))]F(\theta)d\theta. \tag{17}$$

Substituting (17) into (16) yields

$$\begin{aligned} W = \int_{\theta^-}^{\theta^+} \Bigg[& V(Q(p_1(\theta))) - \Big[\Big(\theta - e_1(\theta) + (1-\alpha)\Big(\frac{F(\theta)}{f(\theta)}\Big)\Big)Q(p_1(\theta)) + K + D(e_1(\theta))\Big] \\ & + \beta\Big(V(Q(p_2(\theta))) - \Big[\Big(\theta - e_2(\theta) + (1-\alpha)\Big(\frac{F(\theta)}{f(\theta)}\Big)\Big)Q(p_2(\theta)) + K + D(e_2(\theta))\Big]\Big)\Bigg] f(\theta)d\theta. \end{aligned} \tag{18}$$

The regulator maximizes W with respect to $(p_i(\theta), e_i(\theta)), i = 1,2$, subject to the constraint that the firm chooses its effort optimally. The Lagrangian L formed from the Hamiltonian for the regulator's problem is

$$\begin{aligned} L = \Bigg[& V(Q(p_1(\theta))) - \Big[\Big(\theta - e_1(\theta) + (1-\alpha)\Big(\frac{F(\theta)}{f(\theta)}\Big)\Big)Q(p_1(\theta)) + K + D(e_1(\theta))\Big] \\ & + \beta\Big(V(Q(p_2(\theta))) - \Big[\Big(\theta - e_2(\theta) + (1-\alpha)\Big(\frac{F(\theta)}{f(\theta)}\Big)\Big)Q(p_2(\theta)) + K + D(e_2(\theta))\Big]\Big)\Bigg] f(\theta) \\ & + \lambda_1(\theta)\Big(Q(p_1(\theta)) - D'(e_1(\theta))\Big) + \lambda_2(\theta)\Big(Q(p_2(\theta)) - D'(e_2(\theta))\Big), \end{aligned} \tag{19}$$

where $\lambda_i(\theta)$ is the multiplier associated with the constraint in (9). The necessary optimality conditions thus are[26]

$$p_i(\theta) = \theta - e_i(\theta) + (1-\alpha)\frac{F(\theta)}{f(\theta)} + \frac{\lambda_i(\theta)}{f(\theta)}, \quad i = 1,2, \tag{20}$$

and

$$Q(p_i(\theta)) - D'(e_i(\theta)) - \frac{\lambda_i(\theta)}{f(\theta)}D''(e_i(\theta)) = 0, \quad i = 1,2. \tag{21}$$

Substituting the condition in (9) into (21) yields

$$-\frac{\lambda_i(\theta)}{f(\theta)}D''(e_i(\theta)) = 0, \quad i = 1,2. \tag{22}$$

Since D is strictly convex, the multiplier $\lambda_i(\theta)$ is zero. The necessary condition in (21) is identical to the first-order condition in (9) for the manager's choice of effort, so the regulator can delegate the effort decision to the firm. The effort chosen by the manager is efficient given the quantity the firm produces, but as indicated in (20) the price is distorted from the first-best level so the chosen effort is second-best.

---

[26] This uses the property that $V'(Q(p_i(\theta))) = p_i(\theta)$.

In each period, the price given in (20) equals the marginal production cost $(\theta - e_i(\theta))$ plus the portion $(1-\alpha)$ of the marginal information rents $\frac{F(\theta)}{f(\theta)}$ that the regulator bears. Even though it is implementable, the regulator prefers to deviate from marginal cost pricing because doing so reduces the information rents by more than the reduction in consumer surplus. Consequently, for $\alpha < 1$ the price is greater than marginal cost for all $\theta > \theta^-$. If $\alpha = 1$ the prices are first-best, so the optimal regulatory mechanism maximizes total surplus for every θ. Even though the mechanism is first-best when $\alpha = 1$, politics can be present because the manager earns rents as given in (11). The politics in this case are purely distributive. If $\alpha < 1$, the politics involve both distribution and efficiency.

A sufficient condition for the price in (20) to be optimal is that $\theta + (1-\alpha)\frac{F(\theta)}{f(\theta)}$ is nondecreasing in θ.[27] For many of the commonly-used distributions such as the uniform, normal, and exponential, the term $\frac{F(\theta)}{f(\theta)}$ is increasing in θ. Consequently, for distributions with this property the divergence $\left[p_i(\theta) - (\theta - e_i(\theta))\right]$ of price from marginal cost is an increasing function of θ. In an example to be presented in the next section, the political action is shown to be an increasing function of the divergence of price from marginal cost.

The fixed charges T_i can be recovered from (11) and (7) as

$$T_i(\theta) = \int_{\theta}^{\theta^+} Q(p_i(\theta^o))d\theta^o - p_i(\theta)Q(p_i(\theta)) + (\theta - e_i(\theta))Q(p_i(\theta)) + K + D(e_i(\theta)), \quad i = 1,2. \tag{23}$$

The first term is the incentive payment and the remaining terms are the portion of the firm's costs that are not covered by the revenue $p_i(\theta)Q(p_i(\theta))$.

The effort actually exerted by the manager is determined by (9) and depends on the regulatory instruments only through the quantity $Q(p_i(\theta))$. The marginal product of effort is proportional to that quantity, and since price is greater than marginal cost (unless $\alpha = 1$ or $\theta = \theta^-$), the effort of the firm is lower than the first-best effort. The regulator can, of course, induce the manager to provide the first-best effort but chooses not to do so because it prefers to reduce the rents of the firm by increasing price above marginal cost.

---

[27] This follows from the incentive compatibility requirement.

The properties of this regulatory incentive mechanism are summarized in the following proposition.

Proposition 1: 1) The incentive mechanism specifies prices $p_i(\theta)$ in (20) which equal marginal production cost plus marginal information cost, and prices are the same in both periods. 2) The regulator delegates the effort decision to the firm, and the effort chosen is second-best and less than the first-best effort. 3) The manager earns rents $R_i(\theta)$ in each period, and those rents are greater the more efficient is the firm; i.e., the lower is the parameter θ.

III. Interest Group Competition

Political action to alter the regulatory policy can be directed at either the mandate of the regulatory agency or at the decisions of the agency. For the purposes of this paper the political action will be assumed to be directed at changing the actions of the regulator, so the weight α, which corresponds to the mandate of the regulatory agency, will be assumed to remain unchanged. The political competition thus pertains only to the price that is set in the second period.

The second-period price in (20) is *ex post* inefficient, so there exist Pareto superior pricing policies. For example, the price could be lowered to $p_2(\theta) = \theta$ and the fixed charges could be adjusted to leave the manager with the same return as yielded by the regulatory mechanism. Consumers, however, can do better by reducing both the price and the fixed charges, and it is the competition to do so that is the focus of the political analysis.

The firm in this case does not have an enforceable property right to the rents that can be earned because of its superior information about its costs. Furthermore, this political problem cannot be contracted away because political rights are inalienable, and hence the rightsholders cannot commit themselves not to act on those rights. Indeed, even if it were possible, the firm will be shown not to prefer *ex ante* that the consumer interest group be restrained from political action. *Ex post*, the firm prefers that the consumer interest group not take action, but given the inalienability of political rights, there is no way induce it to refrain from action.

Since the politics in this model result from the political action of the consumer interest group and the firm as applied to the Congress, the regulator is modeled as an ideal bureaucrat who seeks to maximize the *ex ante* weighted surplus specified in its mandate and has the power to commit itself to long-term policies. The regulator can, however, be forced to revise its policy by the Congress. The theory of interaction between interest groups and the Congress and of the interactions between the Congress and the bureaucracy is unsettled. For example, Weingast and Moran (1982)(1983) argue that the bureaucracy is responsive to the preferences of the legislative committee with jurisdiction over it. Moe (1987), however, views the bureaucracy as more insulated from the preferences of committee members and argues that presidential influence is also important. The perspective taken here is that the Congress can affect the regulator's resolve to implement the pricing policy $(p_2(\theta), T_2(\theta)), i = 1, 2$, but the Congress will act only if the pressure on it to do so is sufficiently high.

That pressure results from the political action of the consumer interest group. Given its knowledge of the regulatory mechanism and the observation of the profit of the firm, the interest group can invert the profit function

$$\pi_1(\theta) = R_1(\theta) + D(e_1(\theta)),$$

since

$$\frac{d\pi_1(\theta)}{d\theta} = -Q(p_1(\theta)) + D'(e_1(\theta))e_1'(\theta) < 0,$$

because convexity of $D(\cdot)$ and (9) imply that $e_1'(\theta) < 0$. The interest group thus learns θ from profits.[28] As will be demonstrated in the next section, the profit function in period one remains invertible when political action is anticipated.

Since the focus is on the effect of political action on regulatory policy, that political action will be represented in its simplest form.[29] The consumer interest group is assumed to be able to undertake an "amount" a_I of political action in response to its observation of profit, and to counter that action the firm is assumed to undertake an amount of political action a_F. The political action may represent lobbying effort, campaign contributions, grass-roots activities, etc., and is assumed to determine whether or not the Congress acts to force revision in the regulator's policies. Congressional action need not take the form of legislation but instead may involve threats of legislation to force revisions of policies or direct threats against the regulator such as reductions in budgets or restrictions on its authority.[30] If the Congress forces revision, it had might as well reduce price as far as possible, subject to the firm earning a fair return. Consequently, in this model revision of the regulator's policy results in the first-best price $p_2^*(\theta)$ and fixed charges $T_2^*(\theta) = K + D(e_2^*(\theta))$. If the Congress does not act, the regulator leaves the prices $(p_2(\theta), T_2(\theta))$ in effect. The political action will be assumed to affect the probability that the Congress acts to change the regulatory policy, so let $m(a_I, a_F)$ denote that probability.[31] The function $m(\cdot,\cdot)$ is assumed to be strictly increasing and concave in a_I and strictly decreasing and convex in a_F.

---

[28] It is not necessary that the interest group know the entire regulatory mechanism to be able to infer θ. The interest group can observe profit and revenue ρ_1, so it can compute $\pi_1 - \rho_1 = -\theta Q(p_1(\theta)) - K$. Then, it is straightforward to show that $-\theta Q(p_1(\theta))$ is strictly decreasing in θ, so if the interest group knows $p_1(\theta)$, then θ can be inferred from the observation of profit and revenue. Furthermore, if $p_1(\theta)$ is invertible, θ can be inferred directly from the price.

[29] See Becker (1983) and Denzau and Munger (1986) for models of interest group politics.

The political action and its outcome occur at the beginning of the second period, and the effort of the manager is chosen subsequently. If the political action does not result in a revision in the regulatory pricing policy, the effort of the firm is that characterized in (9). If a revision results, the manager faces a price $p_2^*(\theta) = (\theta - e_2^*(\theta))$ and will thus choose the first-best effort $e_2^*(\theta)$.

Political action is costly for both the consumer interest group and the firm, and those costs are denoted $g(a_I)$ and $h(a_F)$, respectively, where $g(\cdot)$ and $h(\cdot)$ are strictly increasing and convex. The political action costs of the firm are not assumed to be reimbursable by the regulator, so the manager bears the costs.[32]

Upon observation of the profit of the firm, the interest group infers the value of the parameter θ, so the interest group and the firm have common knowledge of θ at the time they engage in political competition. The *ex post* welfare $Y_2(\theta)$ of consumers then is

---

[30] See Ferejohn and Shipan (1988) for an analysis of the effect of threatened action on regulatory behavior.

[31] An alternative model of congressional action would involve an explicit representation of a majority rule equilibrium with the (induced) ideal points of legislators a function of the political action of the interest group and the firm. The formulation used here has the same qualitative properties as this alternative formulation and is convenient for generating examples.

[32] Under conventional revenue-requirements regulation, consumers are typically not required to cover political costs of regulated utilities.

$$\begin{aligned} Y_2(\theta) =& m(a_I,a_F)\Big[V(Q(p_2^*(\theta))) - p_2^*(\theta)Q(p_2^*(\theta)) - T_2^*(\theta)\Big] \\ &+ (1-m(a_I,a_F))\Big[V(Q(p_2(\theta))) - p_2(\theta)Q(p_2(\theta)) - T_2(\theta)\Big] - g(a_I), \end{aligned} \tag{24}$$

which after substitution is

$$\begin{aligned} Y_2(\theta) =& m(a_I,a_F)\Big[V(Q(\theta - e_2^*(\theta))) - (\theta - e_2^*(\theta))Q(\theta - e_2^*(\theta)) - K - D(e_2^*(\theta))\Big] \\ &+ (1-m(a_I,a_F))\Big[V(Q(p_2(\theta))) - (\theta - e_2(\theta))Q(p_2(\theta)) - K - D(e_2(\theta)) - R_2(\theta)\Big] - g(a_I). \end{aligned} \tag{25}$$

The expected return ER_2 to the manager is

$$ER_2 = m(a_I,a_F)\cdot 0 + (1-m(a_I,a_F))R_2(\theta) - h(a_F), \tag{26}$$

where $R_2(\theta)$ is the return if the policy is not revised.

For each θ, a Nash equilibrium $(a_I(\theta), a_F(\theta))$ of the political competition is sought. Given $a_F(\theta)$, the political action of the interest group satisfies

$$m_1(a_I(\theta), a_F(\theta))\Delta Y(\theta) - g'(a_I(\theta)) = 0, \tag{27}$$

where the consumers' stake $\Delta Y(\theta)$ in the political competition is

$$\begin{aligned} \Delta Y(\theta) \equiv& V(Q(\theta - e_2^*(\theta))) - (\theta - e_2^*(\theta))Q(\theta - e_2^*(\theta)) - D(e_2^*(\theta)) \\ &- V(Q(p_2(\theta))) + (\theta - e_2(\theta))Q(p_2(\theta)) + D(e_2(\theta)) + R_2(\theta) \end{aligned} \tag{28}$$

and m_i denotes the partial derivative with respect to the ith argument. Given $a_I(\theta)$, the political action of the firm satisfies

$$-m_2(a_I(\theta), a_F(\theta))R_2(\theta) - h'(a_F(\theta)) = 0, \tag{29}$$

where $R_2(\theta)$ is the firm's stake in the political competition.

To facilitate characteriation of the equilibrium, let $m(a_I,a_F) = \frac{\gamma a_I}{\gamma a_I + a_F}$, $g(a_I) = g \cdot a_I$, and $h(a_F) = h \cdot a_F$. The paramater γ is used to represent an advantage of one of the participants in the political competition with $\gamma > 1$ representing an advantage to the consumer interest group and $\gamma < 1$ representing an advantage to the firm. The equilibrium levels of political action are then

$$a_I(\theta) = \frac{\gamma h R_2(\theta) \Delta Y(\theta)^2}{\left(\gamma h \Delta Y(\theta) + g R_2(\theta)\right)^2}, \tag{30}$$

and

$$a_F(\theta) = \frac{\gamma g R_2(\theta)^2 \Delta Y(\theta)}{\left(\gamma h \Delta Y(\theta) + g R_2(\theta)\right)^2}. \tag{31}$$

The ratio of $a_I(\theta)$ to $a_F(\theta)$ is thus

$$\frac{a_I(\theta)}{a_F(\theta)} = \frac{\Delta Y(\theta)/g}{R_2(\theta)/h}.$$

Consequently, the relative levels of political action depend on the relative ratios of the stake in the political contest to the marginal cost of political action.

The political action of the consumer interest group is a decreasing function of its marginal cost g of political action, and the same is true for the firm. The political action of the interest group is an increasing function of the gain $\Delta Y(\theta)$ from successful political action, since

$$\frac{da_I(\theta)}{d\Delta Y(\theta)} = \frac{a_I(\theta)}{\Delta Y(\theta)} \left[\frac{2 g R_2(\theta)}{\Delta Y(\theta)\left(\gamma h \Delta Y(\theta) + g R_2(\theta)\right)} \right]. \tag{32}$$

Similarly, the political action of the firm is an increasing function of the stake $R_2(\theta)$ in avoiding revision in the regulatory policy. Reducing the price $p_2(\theta)$ reduces $\Delta Y(\theta)$ and increases $R_2(\theta)$, so the regulator can reduce the probability that its policy will be revised by reducing price.

The response to a change of the advantage of the consumer interest group is given by

$$\frac{da_I(\theta)}{d\gamma} = \frac{a_I(\theta)}{\gamma} \left[\frac{g R_2(\theta) - \gamma h \Delta Y(\theta)}{h\left(\gamma h \Delta Y(\theta) + g R_2(\theta)\right)} \right], \tag{33}$$

and similarly,

$$\frac{da_I(\theta)}{dh} = \frac{\gamma}{h} \frac{da_I(\theta)}{d\gamma}.$$

The sign of this derivative depends on the sign of $J \equiv g R_2(\theta) - \gamma h \Delta Y(\theta)$, which may be rewritten as

$$J = R_2(\theta) \gamma h \left[\frac{g}{h\gamma} - \frac{\Delta Y(\theta)}{R_2(\theta)} \right]. \tag{32}$$

Thus the sign of J depends on the difference between the ratio of the marginal costs of political action scaled by the advantage represented by γ and the ratio of the stakes involved. The greater is the stake of the consumer interest group relative to that of the firm or the lower is that interest group's marginal cost of political action, the more likely is J to be negative. The following results can be verified:

$$J>(=)(<)0 \quad \Rightarrow \quad \begin{array}{l} \dfrac{da_I(\theta)}{dR_2(\theta)} <(=)(>)\, 0 \\ \dfrac{da_I(\theta)}{dh} >(=)(<)\, 0. \end{array} \tag{35}$$

$$J>(=)(<)0 \quad \Rightarrow \quad \begin{array}{l} \dfrac{da_F(\theta)}{d\Delta Y(\theta)} >(=)(<)\, 0 \\ \dfrac{da_F(\theta)}{dg} <(=)(>)\, 0. \end{array} \tag{36}$$

The relation between the parameter θ and the political action is complex and will be characterized for the example presented below.

The equilibrium political action results in a probability of revision given by

$$m(a_I(\theta), a_F(\theta)) = \frac{\gamma h \Delta Y(\theta)}{\gamma h \Delta Y(\theta) + g R_2(\theta)}. \tag{37}$$

This probability is an increasing function of h, so the more costly is the political action of the regulator the more likely is revision. Similarly, the more costly is political action to the interest group the less likely is revision. An increase in the competitive advantage, represented by γ, increases the probability of revision. These properties of the political equilibrium confirm the usual intuition underlying interest group competition.

The probability of revision also depends on the parameter θ. The derivative is

$$\frac{dm(a_I(\theta), a_F(\theta))}{d\theta} = \frac{\gamma h g}{\left(\gamma h \Delta Y(\theta) + g R_2(\theta)\right)^2} \left[R_2(\theta) \frac{d\Delta Y(\theta)}{d\theta} - \Delta Y(\theta) \frac{dR_2(\theta)}{d\theta} \right]. \tag{38}$$

For $\alpha = 1$, the price $p_2(\theta) = (\theta - e_2^*(\theta))$ and from (26) $\Delta Y(\theta) = R_2(\theta)$, so the expression in brackets is zero. Consequently, when the politics only involve distribution, the probability of revision is constant in θ.

The effect of the information θ on the probability of revision is given in (38). Using $N(\theta) \equiv \Delta Y(\theta) - R_2(\theta)$, then

$$\frac{d\Delta Y(\theta)}{d\theta} = \frac{dN(\theta)}{d\theta} + \frac{dR_2(\theta)}{d\theta}.$$

For $p_2(\theta) > \theta - e_2(\theta)$, which is the case in equilibrium, and $p_2(\theta)$ is increasing in θ, then $\frac{dN(\theta)}{d\theta} > 0$. Then, letting $O(\theta)$ denote the term in brackets in (38) and substituting for $\frac{d\Delta Y(\theta)}{d\theta}$ yields

$$O(\theta) = R_2(\theta)\frac{dN(\theta)}{d\theta} - N(\theta)\frac{dR_2(\theta)}{d\theta}.$$

Consequently, since $\frac{dR_2(\theta)}{d\theta} < 0$, $O(\theta)$ is positive when $\alpha < 1$. Consequently, when the politics involves both (*ex post*) efficiency and distribution, the probability of revision is a strictly increasing function of the cost parameter θ. That is, the greater are the efficiency stakes, the more likely are the politics to result in a revision of the regulatory policy. Alternatively, the greater is the divergence of price $p_2(\theta)$ from marginal cost the more likely is the regulatory policy to be revised.

These results are summarized qualitatively in the following proposition.

Proposition 2: The equilibrium levels of political action are decreasing functions of their own marginal cost and increasing functions of their own stake in the political action. The cross-effects of costs and stakes in (35) and (36) depend on the sign of J defined in (34). The probability of revision is an increasing function of the consumer interest group's stake, the advantage of that interest group, and the cost of the firm's political action. The probability of revision is a decreasing function of the stake of the firm and of the marginal cost of political action of the interest group. When the politics are purely distributive ($\alpha = 1$), the probability of revision is constant in the cost parameter θ. When the politics involves both distribution and efficiency, the probability of revision is an increasing function of θ.

As an example, let $F(\theta) = \frac{\theta - \theta^-}{\theta^+ - \theta^-}$ on $[\theta^-, \theta^+]$, let $Q(p) = a - bp$, where $a > b(\theta^+ + (1 - \alpha)(\theta^+ - \theta^-))$ and $b < 1$, and let $D(e) = \frac{1}{2}e^2$. For this example, $\frac{F(\theta)}{f(\theta)} = \theta - \theta^-$, so higher values of θ correspond to a greater divergence of price from marginal cost. Then, for the case in which the politics are purely distributive ($\alpha = 1$), the equilibrium, which will be characterized in Section IV and in the Appendix, is[33]

$$p_2^*(\theta) = p_2(\theta) = \frac{\theta - a}{1 - b}, \tag{39}$$

$$e_2^*(\theta) = e_2(\theta) = Q(p_2^*(\theta)) = \frac{a - b\theta}{1 - b}, \tag{40}$$

---

[33] For $\alpha < 1$ the equilibrium cannot be characterized in closed form.

$$\Delta Y(\theta) = R_2(\theta) = \left(\frac{\theta^+ - \theta}{1-b}\right)\left[a - b\left(\frac{\theta^+ + \theta}{2}\right)\right]. \tag{41}$$

When the politics are purely distributive, the equilibrium levels of political action are

$$a_I(\theta) = \frac{\gamma h R_2(\theta)}{(\gamma h + g)^2} \tag{42}$$

and

$$a_F(\theta) = \frac{\gamma g R_2(\theta)}{(\gamma h + g)^2}. \tag{43}$$

The political action of both the interest group and the firm are increasing in the profit of the firm and thus decreasing in the cost parameter θ. The ratio of the levels of political action depends only on the ratio of marginal costs, so the party with the lower marginal cost takes more political action. The equilibrium probability of a revision is then

$$m(a_I(\theta), a_F(\theta)) = \frac{\gamma h}{\gamma h + g}. \tag{44}$$

In the case of purely distributive politics, the equilibrium probability of revision of the regulatory policy is constant in θ, an increasing function of γ and h, and a decreasing function of g. If $\gamma = 1$ so that neither side had a structural advantage and if the marginal costs were the same for both parties, the probability of revision would be one-half.

IV. Political Action and the Design of the Regulatory Incentive Mechanism

The regulatory design presented in Section II and the political response characterized in Section III are brought together in this section as a means of determining how the regulator takes into account the subsequent political action and what effect it has on efficiency and on distribution. The analysis proceeds initially by restricting attention to the class of separating mechanisms which have the property that each pricing policy corresponds to a single value of the cost parameter θ. It is then shown that the regulator always prefers a separating mechanism.

The characterization of the regulatory incentive mechanism when political action is anticipated parallels that in Section II, so only the differences resulting from that political action will be presented. Letting ^{m} denote the case in which revision is possible, the local incentive compatibility condition in (10) becomes

$$\frac{dR^m(\theta)}{d\theta} = -\Big(Q(p_1(\theta)) + \beta\big(m(\theta)Q(p_2^*(\theta)) + (1 - m(\theta))Q(p_2(\theta))\big)\Big), \tag{10a}$$

where $m(\theta) \equiv m(a_I(\theta), a_F(\theta))$, which is an implicit function of the regulatory controls $(p_2(\theta), T_2(\theta))$. The rents $R^m(\theta)$ earned by the firm thus are

$$R^m(\theta) = \int_{\theta}^{\theta^+} \Big[Q(p_1(\theta^o)) + \beta\Big(m(\theta^o)Q(p_2^*(\theta^o)) + (1 - m(\theta^o))Q(p_2(\theta^o))\Big)\Big] d\theta^o. \qquad (11a)$$

The expression corresponding to (15) is then

$$\begin{aligned} W^m = \int_{\theta^-}^{\theta^+} \Big[& V(Q(p_1(\theta))) - p_1(\theta)Q(p_1(\theta)) - T_1(\theta) + \alpha R^m(\theta) \\ & + \beta m(\theta)\Big(V(Q(p_2^*(\theta))) - p_2^*(\theta)Q(p_2^*(\theta)) - T_2^*(\theta)\Big) \\ & + \beta(1 - m(\theta))\Big(V(Q(p_2(\theta))) - p_2(\theta)Q(p_2(\theta)) - T_2(\theta)\Big)\Big] f(\theta)d\theta. \end{aligned} \qquad (15a)$$

Whenever $m(\theta)>0$, the objective function W^m in (15a) is strictly less than W in (15), so the regulator's objective function is lower when the interest group can take political action *ex post.*

The solution to the regulator's problem is characterized in the Appendix and only the results will be summarized here. The price $p_1(\theta)$ is the same as that in (20) for the case in which political action is not present. The price $p_2(\theta)$, however, is the same as the price in (20) only if $\alpha = 1$, and for $\alpha < 1$ the regulator bows to the political pressure of the interest group by lowering the price below that in (20). That price remains above marginal cost, however. As indicated in Section III, the regulator chooses to do so because a lower price reduces the political action of the consumer interest group thus reducing the probability of revision. The regulator thus trades-off the reduction in *ex ante* welfare resulting from a lower price for a higher probability that that price will remain in effect. Since the price is lower than that in (20), the rents of the manager in (11a) are greater than those when political action is not positive. The manager thus prefers *ex ante* that the consumers take political action.

Even when political action is present, the regulator can delegate the effort decisions to the manager, and the effort chosen will be second-best. This results because at the margin the effect of the effort choice on the political action of both the consumer interest group and the firm is proportional to the difference between the marginal product and the marginal disutility of effort, so the choice of effort is not distorted by the political action.

Although consumers are better off *ex post* if their political action results in a revision in the regulatory policy, they are worse off *ex ante* unless $\alpha = 1$ in which case they are equally well off. To see this, let ΔS denote the difference in *ex ante* consumer surplus. Then,

$$\begin{aligned}\Delta S =& W - W^m \\ &+ \alpha \int_{\theta^-}^{\theta^-} \left[\int_{\theta}^{\theta^+} (1-m)\Big((Q(p_2^m(\theta^o)) - Q(p_2(\theta^o))) - m\big(Q(p_2(\theta^o)) - Q_2^*(\theta^o))\big)\Big) d\theta^o \right] f(\theta) d\theta,\end{aligned} \tag{45}$$

where $p_2^m(\theta)$ denotes the optimal price when revision is possible. The integral is nonnegative and is strictly positive for $\alpha<1$. Since $W>W^m$ for $\alpha<1$, consumers are strictly worse off *ex ante* as a result of their political action. If $\alpha = 1$, $\Delta S = 0$.

To implement these prices, it is necessary to make an incentive payment equal to the rents $R^m(\theta)$ in (15*a*), and a period-two payment can only be made in the event that political action is unsuccessful in revising the regulatory policy. The fixed charges $(T_1^m(\theta), T_2^m(\theta))$ that will implement the prices are thus $T_2^m(\theta) = T_2(\theta)$ given in (23) and

$$T_1^m(\theta) = T_1(\theta) + \beta \int_{\theta}^{\theta^+} m(\theta^o) Q(p_2^*(\theta^o)) d\theta^o + \beta \int_{\theta}^{\theta^+} (m(\theta) - m(\theta^o)) Q(p_2^m(\theta^o)) d\theta^o. \tag{46}$$

The next step is to verify that the fixed charges $T_2(\theta)$ in (23) and $T_1^m(\theta)$ in (46) implement those prices in the sense that the firm will choose the pricing policy designed for its cost parameter. After substituting for $T_1^m(\theta)$ and $T_2(\theta)$, the return $R^m(\hat\theta, e_1, e_2^*, e_2; \theta)$ to the firm is

$$\begin{aligned} R^m(\hat\theta, e_1, e_2^*, e_2; \theta) =& (\hat\theta - \theta) Q(p_1(\hat\theta)) - (e_1(\hat\theta) - e_1) Q(p_1(\hat\theta)) + D(e_1(\hat\theta)) - D(e_1) \\ &+ \int_{\hat\theta}^{\theta^+} Q(p_1(\theta^o)) d\theta^o + \beta \int_{\hat\theta}^{\theta^+} m(\theta^o) Q(p_2^*(\theta^o)) d\theta^o \\ &+ \beta m(\hat\theta) \Big[(\hat\theta - \theta) Q(p_2^*(\hat\theta)) - (e_2^*(\hat\theta) - e_2^*) Q(p_2^*(\hat\theta)) + D(e_2^*(\hat\theta)) - D(e_2^*) \Big] \\ &+ \beta(1 - m(\hat\theta)) \Big[(\hat\theta - \theta) Q(p_2^m(\hat\theta)) - (e_2(\hat\theta) - e_2) Q(p_2^m(\hat\theta)) + D(e_2(\hat\theta)) \\ &- D(e_2) + \int_{\hat\theta}^{\theta^+} Q(p_2^m(\theta^o)) d\theta^o \Big] + \beta \int_{\hat\theta}^{\theta^+} (m(\hat\theta) - m(\theta^o)) Q(p_2^m(\theta^o)) d\theta^o. \end{aligned} \tag{47}$$

It is straight-forward to verify that $\hat\theta = \theta$, $e_1 = e_1(\theta)$, $e_2 = e_2(\theta)$, and $e_2^* = e_2^*(\theta)$ satisfy the first-order conditions for the maximization of R^m. A sufficient condition for these to maximize R^m is that

$$A(\theta) \equiv Q(p_1(\theta)) + \beta m(\theta) Q(p_2^*(\theta)) + \beta(1 - m(\theta)) Q(p_2^m(\theta)) \tag{48}$$

is nonincreasing in θ.[34] If $\alpha = 1$ so that the politics are only concerned with distribution, this condition is satisfied, since $m(\theta)$ is constant in θ. When $\alpha < 1$, the probability of revision $m(\theta)$ is increasing in θ, so the sign of the derivative of $A(\theta)$ is ambiguous. Since $m(\theta)$ is multiplied by $\big(Q(p_2^*(\theta)) - Q(p_2^m(\theta))\big)$, however, this condition is likely to be satisfied. For values of θ close to θ^-, $A(\theta)$ is decreasing in θ. If the expression in (48) is increasing in θ on some interval, the approach in Baron and Myerson (1982) must be used to "flatten" the increasing segment through the choice of $p_1(\theta)$ and $p_2(\theta)$.

The final step is to determine if the profit function $\pi_1^m(\theta)$ is strictly decreasing in θ, so that the interest group can invert it to learn θ. The profit in period one is

$$\pi_1^m(\theta) = \int_\theta^{\theta^+} Q(p_1(\theta^o) d\theta^o + \beta \int_\theta^{\theta^+} m(\theta^o) Q(p_2^*(\theta^o)) d\theta^o + \beta \int_\theta^{\theta^+} \big(m(\theta) - m(\theta^o)\big) Q(p_2^m(\theta^o)) d\theta^o. \tag{49}$$

The derivative is

$$\frac{d\pi_1^m(\theta)}{d\theta} = -Q(p_1(\theta)) - \beta m(\theta) Q(p_2^*(\theta)) + \beta m'(\theta) \int_\theta^{\theta^+} Q(p_2^m(\theta^o)) d\theta^o. \tag{50}$$

Since $m'(\theta)$ is nonegative, the sign of this derivative is ambiguous, although for α close to one it is negative.

The following proposition summarizes the results:

Proposition 4: 1) The price $p_1(\theta)$ is unaffected by the anticipation of political action, but for $\alpha <(=) 1$ the period-two price $p_2^m(\theta)$ is lower than (equal to) the price in the absence of political action. 2) The regulator delegates the effort decision to the firm and the resulting effort is second-best. 3) The rents earned by the manager are greater when political action is possible than when it is not.[35] Consequently, the manager prefers that the consumer interest group be unable to commit not to take political action *ex post*. The firm prefers to commit itself *ex ante* not to take political action, but *ex post* it prefers to counter the political action of the consumer interest group. 4) From an *ex ante* perspective, for $\alpha < (=) 1$ consumers are worse (equally well) off as a consequence of their inability to commit not to take political action *ex post*. Consumers thus prefer to be able to commit credibly not to take political action, but because political rights are inalienable, they are unable to do so.

---

[34] See Baron and Myerson (1982).

[35] To see this, note that in the expression for the rents in (10a) $Q(p_2^*(\theta)) > Q(p_2^m(\theta))$ $\forall \theta > \theta^-$, so R^m is an increasing function of the probability $m(\theta)$ of revision.

The analysis in this section is based on the assumption that the regulatory incentive mechanism is separating so that a different pricing policy is associated with each possible value of the cost parameter. The regulator could choose a pooling mechanism that sets the same price for a set of values of the cost parameter. For example, the price $\bar{p}_1$ could be established for all θ. The interest group, however, can observe the profit of the firm in the first period, and that profit is a function of θ given by

$$\pi_1(\theta) = \int_{\theta}^{\theta^+} Q(\bar{p}_1)d\theta^o + D(\bar{e}_1(\theta)) = (\theta^+ - \theta)Q(\bar{p}_1) + D(\bar{e}_1(\theta)),$$

where $\bar{e}_1(\theta)$ is the effort given $Q(\bar{p}_1)$. Observation of the profit and knowledge of the regulatory mechanism allows the interest group to determine the cost parameter θ, and hence the political action analogous to that characterized in Section III can be anticipated. The regulator, of course, can do better in terms of the *ex ante* surplus by choosing the separating mechanism characterized in the Appendix.

V. Extensions and Generalizations

Political competition has been represented above as having a dichotomous outcome of either the prices specified by the regulatory mechanism or the first-best prices with the political action affecting the probability of revision. An alternative specification would be to model the period-two price as a direct function of the political action of the interest group and of the firm. In such a formulation, the regulator would have no reason to specify a price for the second-period because the price would be determined directly by the political action.

An alternative characterization of the political competition would involve a model of the relation between the distributive consequences of a regulatory incentive mechanism and the preferences of legislators. In the simplest model, the alternative set could be specified as the set of prices $[p_2^*(\theta), p_2(\theta)]$, and the preferences of the median legislator could be specified as depending on the political action of the consumer interest group and the firm.

The revised price then would result from the effect of that action on the ideal point of the median legislator. The qualitative properties of such an equilibrium would be the same as those characterized in Section III.

VI. Conclusions

The distributive consequences of regulatory policies create incentives for political action, and regulators and policy-makers are not well-insulated from that action. In particular, if a regulatory innovation either results in adverse distributive consequences for some interest group or creates an opportunity to seek economic profits or rents that would result as a consequence of improved efficiency, political action can be expected. The initiative for political action may be taken by interest groups or by political entrepreneurs who seek to organize and to lead as a means of representing, or catering to, organized and unorganized interests. Such political action is not always successful, in part because opposition is also created by distributive consequences, but often it forces revision in or adaptation of regulatory policies. When political action is possible, policy-makers and designers of incentive mechanisms must for at least two reasons be more than cognizant of the political reality. First, a regulatory incentive mechanism will not lead to the anticipated performance if those affected by the mechanism have incentives to take political action in addition to making their market transactions. Second, by anticipating political action and designing mechanisms that take into account that action, regulators can improve the chances that improved efficiency will result.

The model considered here is intended to initiate the study of the design of regulatory incentive mechanisms when political action can be anticipated. The model considers only one type of incentive problem resulting from asymmetric information about costs. Regulation, however, involves a wide variety of incentive problems, and the politics associated with those problems may not be identical to that analyzed here. Furthermore, the model incorporates only *ex post* political action, but politics may be present at the *ex ante* or design stage as well. The model also abstracts away from institutional characteristics that can be important in determining whether political action will be successful in altering regulatory policies. More general treatments of political action and the design of regulatory policies, however, await further research.

APPENDIX

Optimal control theory will be used to characterize the solution to the regulator's problem when political action is present. Let $R^m(\theta)$ be the state variable defined by

$$\begin{aligned} R^m(\theta) &= p_1(\theta)Q(p_1(\theta)) + T_1(\theta) - (\theta - e_1)Q(p_1(\theta)) - K - D(e_1) \\ &\quad + \beta\bar{m}(p_2(\theta),T_2(\theta))\Big(p_2^*(\theta)Q(p_2^*(\theta)) + T_2^*(\theta) - (\theta - e_2^*)Q(p_2^*(\theta)) - K - D(e_2^*)\Big) \\ &\quad + \beta\big(1 - \bar{m}(p_2(\theta),T_2(\theta))\big)\Big(p_2(\theta)Q(p_2(\theta)) + T_2(\theta) - (\theta - e_2)Q(p_2(\theta)) - K - D(e_2)\Big), \end{aligned} \tag{A.1}$$

and as in (10)

$$R^{m\prime}(\theta) = -Q(p_1(\theta)) - \beta\bar{m}(p_2(\theta),T_2(\theta))Q(p_2^*(\theta)) - \beta(1 - \bar{m}(p_2(\theta),T_2(\theta)))Q(p_2(\theta)), \tag{A.2}$$

where $\bar{m}(p_2(\theta),T_2(\theta))$ is the probability of revision expressed in terms of the controls $p_2(\theta)$ and $T_2(\theta)$.

The objective function W^m of the regulator is expressed as

$$\begin{aligned} W^m = \int_{\theta^-}^{\theta^+} \Big[& V(Q(p_1(\theta))) - p_1(\theta)Q(p_1(\theta)) - T_1(\theta) + \alpha R^m(\theta) \\ & + \beta\bar{m}(p_2(\theta),T_2(\theta))\Big(V(Q(p_2^*(\theta))) - p_2^*(\theta)Q(p_2^*(\theta)) - T_2^*(\theta)\Big) \\ & + \beta(1 - \bar{m}(p_2(\theta),T_2(\theta)))\Big(V(Q(p_2(\theta))) - p_2(\theta)Q(p_2(\theta)) - T_2(\theta)\Big)\Big] f(\theta)d\theta. \end{aligned} \tag{A.3}$$

The Lagrangian $L(\theta)$ formed from the Hamiltonian for the regulator's program, ignoring initially the constraints on the return to the manager in each period, is

$$\begin{aligned}
L(\theta) = \Big[& V(Q(p_1(\theta))) - p_1(\theta)Q(p_1(\theta)) - T_1(\theta) + \alpha R^m(\theta) \\
& + \beta\bar{m}(p_2(\theta), T_2(\theta))\Big(V(Q(p_2^*(\theta))) - p_2^*(\theta)Q(p_2(\theta)) - T_2^*(\theta)\Big) \\
& + \beta(1 - \bar{m}(p_2(\theta), T_2(\theta)))\Big(V(Q(p_2(\theta))) - p_2(\theta)Q(p_2(\theta)) - T_2(\theta)\Big)\Big] f(\theta) \\
& + \phi(\theta)\Big[p_1(\theta)Q(p_1(\theta)) + T_1(\theta) - (\theta - e_1)Q(p_1(\theta)) - K - D(e_1) \\
& + \beta\bar{m}(p_2(\theta), T_2(\theta))\Big(p_2^*(\theta)Q(p_2(\theta)) + T_2^*(\theta) - (\theta - e_2^*)Q(p_2^*(\theta)) - K - D(e_2^*)\Big) \\
& + \beta(1 - \bar{m}(p_2(\theta), T_2(\theta)))\Big(p_2(\theta)Q(p_2(\theta)) + T_2(\theta) - (\theta - e_2)Q(p_2(\theta)) - K \\
& - D(e_2)\Big) - R^m(\theta)\Big] + \lambda_1(\theta)(Q(p_1(\theta)) - D'(e_1)) + \beta\lambda_2^*(\theta)(Q(p_2^*(\theta)) - D'(e_2^*)) \\
& + \beta\lambda_2(\theta)(Q(p_2(\theta)) - D'(e_2)) \\
& - \mu(\theta)\Big[Q(p_1(\theta)) + \beta\bar{m}(p_2(\theta), T_2(\theta))Q(p_2^*(\theta)) + \beta(1 - \bar{m}(p_2(\theta), T_2(\theta)))Q(p_2(\theta))\Big],
\end{aligned} \tag{A.4}$$

where $\phi(\theta)$ is the multiplier associated with the definition of the state variable, $\lambda_i(\theta), i = 1, 2$, and $\lambda_2^*(\theta)$ are the multipliers associated with the choices of effort, and $\mu(\theta)$ is the costate variable.

The necessary optimality conditions are

$$\begin{aligned}
\frac{\partial L(\theta)}{\partial p_1(\theta)} = & - Q(p_1(\theta))f(\theta) + \phi(\theta)\Big[Q(p_1(\theta)) - p_1(\theta)Q'(p_1(\theta)) - (\theta - e_1(\theta))Q'(p_1(\theta))\Big] \\
& + \lambda_1(\theta)Q'(p_1(\theta)) - \mu(\theta)Q'(p_1(\theta)) = 0
\end{aligned} \tag{A.5}$$

$$\begin{aligned}
\frac{\partial L(\theta)}{\partial p_2(\theta)} = & - \beta(1 - \bar{m}(p_2(\theta), T_2(\theta)))Q(p_2(\theta))f(\theta) \\
& + \beta(1 - \bar{m}(p_2(\theta), T_2(\theta)))\Big[Q(p_2(\theta)) + p_2(\theta)Q'(p_2(\theta)) - (\theta - e_2(\theta))Q'(p_2(\theta))\Big] \\
& - \mu(\theta)\beta(1 - \bar{m}(p_2(\theta), T_2(\theta)))Q'(p_2(\theta)) \\
& + \beta\bar{m}_{p_2}\Big[\Delta W f(\theta) - \mu(\theta)\big(Q(p_2^*(\theta)) - Q(p_2(\theta))\big)\Big] = 0
\end{aligned} \tag{A.6}$$

$$\frac{\partial L(\theta)}{\partial e_1(\theta)} = \phi(\theta)\big(Q(p_1(\theta)) - D'(e_1(\theta))\big) - \lambda_1(\theta)D''(e_1(\theta)) = 0 \tag{A.7}$$

$$\begin{aligned}\frac{\partial L(\theta)}{\partial e_2(\theta)} =&\beta\phi(\theta)(1-\bar{m}(p_2(\theta),T_2(\theta)))\big(Q(p_2(\theta))-D'(e_2(\theta))\big)-\lambda_2(\theta)D''(e_2(\theta))\\ &+\beta\bar{m}_{e_2}\Big[\Delta W f(\theta)-\mu(\theta)\big(Q(p_2^*(\theta))-Q(p_2^*(\theta))\big)\Big]=0\end{aligned} \tag{A.8}$$

$$\begin{aligned}\frac{\partial L(\theta)}{\partial e_2^*(\theta)} =&\phi(\theta)\bar{m}(p_2(\theta),T_2(\theta))\big(Q(p_2^*(\theta))-D'(e_2^*(\theta))\big)-\lambda_2^*(\theta)D''(e_2^*(\theta))\\ &+\beta\bar{m}_{e_2^*}\Big[\Delta W f(\theta)-\mu(\theta)\big(Q(p_2^*(\theta))-Q(p_2(\theta))\big)\Big]=0\end{aligned} \tag{A.9}$$

$$\frac{\partial L(\theta)}{dT_1(\theta)} = -f(\theta)+\phi(\theta)=0 \tag{A.10}$$

$$\begin{aligned}\frac{\partial L(\theta)}{\partial T_2(\theta)} =&\beta(1-\bar{m}(p_2(\theta),T_2(\theta)))(-f(\theta)+\phi(\theta))\\ &+\beta\bar{m}_{T_2}\Big[\Delta W f(\theta)-\mu(\theta)\big(Q(p_2^*(\theta))-Q(p_2^*(\theta))\big)\Big]=0\end{aligned} \tag{A.11}$$

$$\frac{\partial L(\theta)}{\partial R^m(\theta)} = -\mu'(\theta)\rightarrow -\mu'(\theta)=\alpha f(\theta)-\phi(\theta) \tag{A.12}$$

where $\bar{m}_x$ denotes the partial derivative with respect to x and

$$\begin{aligned}\Delta W(\theta) \equiv& V(Q(p_2^*(\theta)))-(\theta-e_2^*(\theta))Q(p_2^*(\theta))-D(e_2^*(\theta))\\ &-\Big(V(Q(p_2(\theta)))-(\theta-e_2(\theta))Q(p_2(\theta))-D(e_2(\theta))\Big).\end{aligned} \tag{A.13}$$

Since $(p_2^*(\theta),e_2^*(\theta))$ maximizes $[V(Q(p))-(\theta-e)Q(p)-D(e)]$, $\Delta W(\theta)$ is nonnegative and is strictly positive for any $p_2(\theta)\neq p_2^*(\theta)$.

From $(A.10)$, $\phi(\theta)=f(\theta)$, so from $(A.12)$

$$\mu(\theta)=(1-\alpha)F(\theta). \tag{A.14}$$

Using $(A.14)$,

$$\begin{aligned}\Delta Z(\theta) &\equiv \Delta W(\theta)f(\theta)-\mu(\theta)(Q(p_2^*(\theta))-Q(p_2(\theta)))\\ &= V(Q(p_2^*(\theta)))-\left(\theta-e_2^*(\theta)+(1-\alpha)\frac{F(\theta)}{f(\theta)}\right)Q(p_2^*(\theta))-D(e_2^*(\theta))\\ &-V(Q(p_2(\theta)))+\left(\theta-e_2(\theta)+(1-\alpha)\frac{F(\theta)}{f(\theta)}\right)Q(p_2(\theta))+D(e_2(\theta)).\end{aligned} \tag{A.15}$$

The probability of revision function is from (37)

$$\bar{m}(p_2(\theta), T_2(\theta)) = \frac{\Delta Y(\theta)}{\Delta Y(\theta) + \eta R_2(\theta)}, \tag{A.16}$$

where $\eta \equiv \frac{\rho}{\gamma h}$ and

$$\begin{aligned} \Delta Y(\theta) =& V(Q(p_2^*(\theta))) - (\theta - e_2^*(\theta))Q(p_2^*(\theta)) - D(e_2^*(\theta)) - K \\ & - V(Q(p_2(\theta))) + p_2(\theta)Q(p_2(\theta)) + T_2(\theta) \end{aligned} \tag{A.17}$$

and

$$R_2(\theta) = p_2(\theta)Q(p_2(\theta)) + T_2(\theta) - (\theta - e_2)Q(p_2(\theta)) - K - D(e_2). \tag{A.18}$$

For any prices that leave $R_2(\theta) \geq 0$, $\Delta Y(\theta)$ is nonnegative and if $p_2(\theta) \neq p_2^*(\theta)$, $\Delta Y(\theta) > 0$.

The following derivatives can then be verified:

$$\frac{\partial \bar{m}}{\partial T_2(\theta)} = -\frac{\eta \Delta W}{\left(\Delta Y(\theta) + \eta R_2(\theta)\right)^2} \tag{A.19}$$

$$\frac{\partial \bar{m}}{\partial e_2(\theta)} = -\frac{\eta \Delta Y\left(Q(p_2(\theta)) - D'(e_2)\right)}{\left(\Delta Y(\theta) + \eta R_2(\theta)\right)^2} \tag{A.20}$$

$$\frac{\partial \bar{m}}{\partial p_2(\theta)} = \frac{\eta\left(Q(p_2(\theta))\Delta W - \Delta Y\left(p_2(\theta) - (\theta - e_2)\right)Q'(p_2(\theta))\right)}{\left(\Delta Y(\theta) + \eta R_2(\theta)\right)^2} \tag{A.21}$$

Since $\Delta W(\theta) \geq 0$, the probability of revision is a nonincreasing function of $T_2(\theta)$ and is a strictly decreasing function if $p_2(\theta) \neq p_2^*(\theta)$. Consequently, an increase in the fixed charges specified in the regulatory mechanism makes the probability of revision smaller.

The derivative $\frac{\partial \bar{m}}{\partial e_2(\theta)}$ in $(A.20)$ is proportional to $\left(Q(p_2(\theta)) - D'(e_2)\right)$. From (9) this equals zero at the effort chosen by the manager. Then, in the adjoint equation $(A.8)$ the multiplier $\lambda_2(\theta) = 0$. It also follows that $\lambda_1(\theta) = 0$ and $\lambda_2^*(\theta) = 0$. The regulator thus continues to want to delegate the effort decisions to the firm.

The price $p_1(\theta)$ for period one can be determined from $(A.10)$, $(A.14)$, and $(A.5)$ and is the same as that characterized in (20) in the absence of political action. Assuming that the Lagrangian is concave in $p_2(\theta)$. To characterize the period-two price $p_2(\theta)$, suppose that that price were the same as that in (20). Then, $(A.6)$ evaluated at that price is

$$\frac{\partial L(\theta)}{\partial p_2(\theta)} = \beta \bar{m}_{p_2} \Delta Z. \tag{A.22}$$

From $(A.15)$ $\Delta Z(\theta)<0$ since $\Delta Z(\theta)$ is maximized at that price. Since $\Delta Y(\theta)>0$, then from $(A.21)$ $\frac{\partial \bar{m}}{\partial p_2(\theta)}$ is positive. The derivative in $(A.22)$ is thus negative. The price $p_2(\theta)$ is thus lower than the price when political action is not possible. Next, suppose that $p_2(\theta) = p_2^*(\theta)$, so $e_2(\theta) = e_2^*(\theta)$. Then, $\Delta W(\theta) = 0$, so $\frac{\partial \bar{m}}{\partial p_2(\theta)} = 0$. The derivative in $(A.6)$ then is positive. The optimal price $p_2(\theta)$ thus is greater than the first-best price and is lower than the price when political action is not present.

The period-two price $p_2(\theta)$ is thus such that $\Delta W(\theta)>0$, so from $(A.19)$ the probability of revision is decreasing in the the fixed charges $T_2(\theta)$. Then, since $\Delta Z(\theta)<0$, using $(A.10)$ in $(A.11)$ implies that $\frac{\partial L(\theta)}{\partial T_2(\theta)}>0$. The regulator thus prefers high fixed charges in period two which would be covered by low fixed charges in the first period.

To determine the optimal fixed charges $T_2(\theta)$, note that the fact that the return of the manager must be nonnegative for all θ implies that

$$T_2^m(\theta) = \int_{\theta}^{\theta^+} Q(p_i(\theta^o))d\theta^o - p_i(\theta)Q(p_i(\theta)) + (\theta - e_i(\theta))Q(p_i(\theta)) + K + D(e_i(\theta)) + \bar{T},$$

where $\bar{T}$ is a nonnegative constant. The requirement that the return of the manager be nonnegative in period one implies that

$$\bar{T} = \frac{R_2^m(\theta^+)}{\beta(1-\bar{m})}.$$

Consequently, if the regulator prefers to increase the stake of the firm above $R_2^m(\theta) = \int_{\theta}^{\theta^+} Q(p_2(\theta^o))d\theta^o$, it is necessary to shift upwards the return function $R_2^m(\theta)$ by $R_2^m(\theta^+)$. The regulator will prefer to do so if the political return is such that it reduces the probability of revision by an amount that offsets the cost of $(1-\alpha)$. Upon simplification, the relevant derivative is

$$\frac{\partial \int L(\theta)d\theta}{\partial \bar{T}} = -(1-\alpha) + \int_{\theta^-}^{\theta^+} \left[\frac{\bar{m}_{T_2}}{1-\bar{m}} \Delta Z(\theta)\right] f(\theta)d\theta. \qquad (A.23)$$

At $\alpha = 1$, the terms $\bar{m}_{T_2}$ and $\Delta Z(\theta)$ equal zero, so the derivative in $(A.23)$ equals zero. The second derivative is positive, so $R^m(\theta^+)$ either equals zero or is unbounded. Although no conclusion can be reached at this level of generality, it seems reasonable that $R^m(\theta^+)$ equals zero. In this case, the return $R_2^m(\theta)$ in period two is $R_2^m(\theta) = \int_{\theta}^{\theta^+} Q(p_2(\theta^o))d\theta^o$.

REFERENCES

Baron, David P. (1987), "The Design of Regulatory Mechanisms and Institutions," in *The Handbook of Industrial Organization*, R. Schmalansee and R. Willig, eds., forthcoming.

Baron, David P. (1988a) "Information, Incentives, and Commitment in Regulatory Mechanisms: Regulatory Innovation in Telecommunications," Working Paper, Stanford University, February.

Baron, David P. (1988b) "Regulation and Legislative Choice," *Rand Journal of Economics*, (forthcoming).

Baron, David P. and David Besanko (1984), "Regulation and Information in a Continuing Relationship," *Information and Economics and Policy*, Vol. 1, pp. 267-301.

Baron, David P. and David Besanko (1987a), "Monitoring, moral hazard, asymmetric information, and risk sharing in procurement contracting," *Rand Journal of Economics*, 18 (Winter), 509-532.

Baron, David P. and David Besanko (1987b), "Commitment and Fairness in a Dynamic Regulatory Relationship," *Review of Economic Studies*, Vol. 54, pp. 413-436.

Baron, David P. and Roger B. Myerson (1982), "Regulating a Monopolist with Unknown Costs," *Econometrica*, Vol. 50, pp. 911-930.

Besanko, David and David E.M. Sappington (1987), "Designing Regulatory Policy with Limited Information." In *Fundamentals of Pure and Applied Economics: Government Ownership and Regulation of Economic Activity*, Elizabeth Bailey, ed.

Bhattacharyya, S.K. and Laughhunn, D.J. (1987), "Price Cap Regulation: Can We Learn From the British Telecom Experience?", *Public Utilities Fortnightly*, October 15, 1987, 22-29.

Becker, Gary S. (1983), "A Theory of Competition among Pressure Groups for Political Influence," *Quarterly Journal of Economics*, 98 (August), 371-400.

Caillaud, B., R. Guesnerie, P. Rey, and J. Tirole (1988), "Government intervention in production and incentives theory: a review of recent contributions," *RAND Journal of Economics*, 19 (Spring), 1-26.

Denzau, Arthur T. and Michael C. Munger (1986), "Legislators and Interest Groups: How Unorganized Interests Get Represented," *American Political Science Review*, 80 (March), 89-106.

Ferejohn, John and Charles Shipan (1988), "Congressional Influence on Telecommunications Policy," Working Paper, Stanford University.

Helman, Leonard A. (1984), "Rulemaking in Lieu of Ad Hoc Case-by-Case Adjudication: Automatic Adjustment Clauses, the New Mexico Experiment," in *Regulatory Reform*, J.R. Foster, G.R. Hall, S.R. Holmberg, C.F. Phillips, Jr., and R.L. Wallace, eds., Institute for Study of Regulation, Washington, D.C., pp. 148-153.

Laffont, Jean-Jacques and Jean Tirole (1986), "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614-641.

Laffont, Jean-Jacques and Jean Tirole (1987), "Comparative Statics of the Optimal Dynamic Incentive Contract," *European Economic Review*, 31, 901-926.

Laffont, Jean-Jacques and Jean Tirole (1988), "The Dynamics of Incentive Contracts," *Econometrica* (forthcoming).

Moe, Terry M. (1987), "An Assessment of the Positive Theory of 'Congressional Dominance'", *Legislative Studies Quarterly*, 12 (November), 475-520.

Noll, Roger E. (1987), "Economic Perspectives on the Politics of Regulation," in *Handbook of Industrial Organization*, Richard Schmalensee and Robert Willig (eds.), North-Holland, (forthcoming).

OFTEL (1988),"The Regulation of British Telecom's Prices," Director General of Telecommunications, London, (January).

Peltzman, Sam (1976), "Toward a More General Theory of Regulation," *Journal of Law and Economics*, 19, 211-240.

Posner, Richard A. (1974), "Theories of Economic Regulation," *Bell Journal of Economics*, 5, 335-358.

Romer, Thomas and Howard Rosenthal (1986), "Modern Political Economy and the Study of Regulation," Working Paper, Carnegie-Mellon University.

Salant, David J. and Woroch, Glenn A. (1988), "Trigger Price Regulation," GTE Laboratories, Waltham, MA, (July).

Sappington, David E.M. (1983), "Optimal Regulation of a Multiproduct Monopoly with Unknown Technological Capabilities," *Bell Journal of Economics*, 14, 453-463.

Sappington, David E.M. and Joseph E. Stiglitz (1986), "Information and Regulation," Working Paper, Bellcore.

Sibley, David (1988), "Asymmetric Information, Incentives and Ceiling Price Regulation," Working Paper, Bellcore.

Stigler, George J. (1971), "The Theory of Economic Regulation," *Bell Journal of Economics*, 2, 3-19.

Vickers, John and George Yarrow (1988), *Privatization: An Economic Analysis*, MIT Press, Cambridge, MA.

Vogelsang, Ingo (1988), "Price Cap Regulation of telecommunications Services: A Long-Run Approach," RAND, Santa Monica, CA, (February).

Weingast, Barry R. and Mark J. Moran (1982), "The Myth of the Runaway Bureaucracy," *Regulation*, 6, 33-38.

Weingast, Barry R. and Mark J. Moran (1983), "Bureaucratic Discretion or Congressional Control? Regulatory Policy-making by the Federal Trade Commission," *Journal of Political Economy*, 91, 765-780.

Preliminary Draft

Efficiency Issues in Competitive Allocation of the Broadcast Spectrum

by

Severin Borenstein*

Prepared for presentation at the Sixteenth Annual Research Conference on Telecommunications Policy, Airlie House, Airlie, VA, October 30 - November 1, 1988.

---

*Department of Economics and Institute of Public Policy Studies, University of Michigan, Ann Arbor, MI 48109-1220, 313-764-6133. Janet Netz and Neil Roseman provided excellent research assistance. Financial support from the National Science Foundation is gratefully acknowledged. I alone am responsible for any errors that remain.

I. Introduction

For more than half a century, the Federal Communications Commission (FCC) has exercised its responsibility for determining who will be allowed to broadcast over a wide frequency spectrum. For many years, technological constraints and a dearth of broadcasters kept this problem relatively simple. Radio broadcasting were few in number and operated only within certain bandwidths that were not valuable for many other uses. Likewise, when television came into being, it too made use of frequencies that had very low or zero opportunity costs. Thus, the FCC's task was usually one of assuring that stations did not broadcast materials considered harmful to society, rather than determining which applicant would most greatly benefit society.

As these media have grown, so have the number and demands of alternative users of the frequency spectrum. Though the FCC at first saw almost no opportunity cost to license assignment, such costs are now substantial. Alternative uses of the frequency spectrum, both intramedia and intermedia, must now be analyzed.[1] Consistent with the regulatory reform movement of the late 1970's and 1980's, many industry observers and regulators of spectrum use have suggested that a market allocation of spectrum rights – usually by auction, nearly always permitting resale – would be an improvement over the administrative approach that the FCC had adopted. Some have even claimed that market allocation would achieve first-best allocation in most cases, *i.e.*, each segment of the broadcast spectrum would be put to the use in which it generated the greatest social value.[2] Though this notion of economic efficiency through market allocation of spectrum rights is intuitively appealing, unfortunately it is not generally correct.

This paper explains why market allocation of operating licenses does not necessarily lead to their efficient allocation – a result that is explored in greater detail in Borenstein (1988) – and then suggests a route for applying these considerations to the allocation of spectrum rights. The failure of competitive license allocation to achieve first-best efficiency occurs because the private value of a license to its purchaser reflects only the profits the firm can then earn with the license, while the full social value includes both the producer profits and the consumer benefits from the license use. A competitive market for licenses will, of course, allocate them to the firms that will generate the greatest profits from their use. Though total social benefits are often positively correlated with profits of the license holder, unless that correlation is perfect, a ranking of license uses by the social value they generate can deviate from a ranking by profitability.

---

[1] Geller (1985) and others have referred to the intramedia/intermedia distinction as being separate problems of spectrum allocation and spectrum assignment.

[2] See Geller (1985) or Kwerel and Felker (1985).

One task then of a policymaker considering a market approach to spectrum allocation should be to analyze the degree to which expected profits of each potential license holder reflect the change in total social surplus (consumer surplus plus profits of all producers in the market) that it is likely to generate. After reviewing the reasons why a competitive market for any kind of operating licenses may fail to attain first-best allocation, I discuss both approaches to and limits on the regulator's analysis of the correlation between private and social value of license use. I conclude that the prospects for improving upon a market allocation of spectrum rights is not great in most cases, but that further studies of the efficiency of market outcomes, perhaps while market allocation is in use, may allow the government to avoid some of the most serious pitfalls.

II. Potential Inefficiencies in Competitive Markets for Operating Licenses

The private value of an operating license will diverge from the social value of its use for two reasons. First, the license holder's private valuation will not reflect the consumer surplus from its licensed output. The degree of this understatement of social surplus will differ across potential uses of a license, possibly causing the ranking of uses by social surplus to differ from a ranking by private profits. Second, some of the profit that the new licensee generates may not be new surplus, but rather a transfer of profits from incumbent firms in the market that it chooses to enter with the license. The new licensee's profits will then overstate the increase in industry profits due to its operations. The degree of this overstatement of the change in profits will differ across potential uses of the license, again making it possible that a ranking of uses by resulting social surplus will differ from a ranking by profits of the new licensee.

The first effect is probably the more intuitive and is also the more easily demonstrated. Consider two monopolists in different markets, both with constant marginal cost equal to 1. The first faces a demand of $Q_1 = 20 - P_1$ and the second faces a demand of $Q_2 = 200/P_2^2$. Solving for the profit maxima reveals that $Q_1^m = 9.5$, leading to a profit of \$90.25 in market 1 and total surplus of \$135.38. In market 2, however, $Q_2^m = 50$, with a profit of \$50 and total surplus of \$150. Market 2 generates lower profits, but the social value of a single monopoly license would be greater if used in market 2 than in market 1. Differences in the shapes of the demand curves that the two firms face lead them to capture different proportions of the total surplus that their monopoly operations generate.

The second effect is closely related to the concept of rent-seeking behavior. The private returns that accrue to a firm from its licensed production may be greater than the change in social surplus due to that firm's output in the market. If a firm uses a license to enter a market in which other profit-making firms already operate, then while the entrant may earn profits (due, for instance, to imperfect competition among the firms in the market), it will also lower the profits of the incumbents. In fact, some of the private value of entry may be just the prospect of profit transfers from the incumbent firms to the entrant. Such transfers have no net social value, so the potential entrant will tend to overvalue the license relative to the social value of its operations.

Consider, for instance, two applicants for a radio broadcasting license. The first will offer programming very similar to that offered by four incumbent stations in the market. Assume that a fixed (or highly inelastic with respect to the number of stations) quantity of consumers listen to this format and that they distribute themselves evenly among stations offering the format. If assigned the license, the first applicant would then attract one-fifth of those listeners who like the format. If the format is popular, the applicant's expected profits from replicating the programming of four existing stations could still be quite large. Its willingness-to-pay for a broadcasting license could quite possibly be larger than that of a second applicant who would offer a new and unique format that would attract fewer listeners, but would mostly attract people who otherwise would not have listened to the radio. The former station would demonstrate excessive valuation of a license due to a "business stealing" effect. Some of its profits would be transfers from incumbents in the market. Of course, consumers might still benefit from the programming of the first station, because it would make greater variety available within the comparatively crowded format. It seems likely, however, that this contribution to social welfare would be smaller *relative* to the applicant's expected profits than would be the case with the second license applicant.[3]

The two reasons that private valuation of a license is likely to differ from the social value of its use – and thus may be misallocated by a competitive market for licenses – stem from two quite distinct facets of licenses and the uses to which they might be put. First, private valuation of an input, such as a license, will generally understate social valuation (and by differing proportions in different uses) when consumer surplus created by the possible uses of the input is significant, but is ignored in the allocation process. This will be the case anytime an input is "lumpy," *i.e.*, allocation of the marginal (but lumpy) unit of input creates inframarginal units of output. These inframarginal units of output, unlike the marginal unit of output in a market, create first-order consumer surplus. Thus, when the indivisible units of an input each create, or are associated with the production of, substantial units of output, non-trivial consumer surplus will be created by marginal input allocation decisions, but will be ignored by the private market. Clearly, an operating license, such as the exclusive right to broadcast over a bandwidth of the frequency spectrum, is such a lumpy input.

Second, private valuation of an input will tend to overstate social value in a given use when the private returns stem in part from transfers of surplus (profits) from incumbent sellers in the market. This can occur only when economic rents would be earned in a market in the absence of the application of the input unit to that market. That is, there must be economic rents for the new licensee to capture.[4] It is important to distinguish such economic rents from accounting profits attributable to scarcity rents, particularly those due to the incumbents' possession of an operating right made scarce by nature or technology, not by administrative, political, or governmental restraints.[5]

---

[3] This argument was first advanced by Steiner (1952) and also presented by Spence and Owen (1977). Borenstein (1986) offers a more rigorous model in which a license can be used to enter any one of many spatially competitive markets. It is demonstrated that as a market becomes more crowded, the ratio of the new entrant's profits to the increase in total surplus from its operations rises. Borenstein (1988) derives similar results for markets in which competition is Nash in quantities.

The second effect, which causes the private valuation of a license to overstate its social value, will persist even if the input can be allocated (and used) in infinitessimal quantities, though the first effect, which causes the private valuation of a license to overstate its social value, would cease. The first effect would persist even if competition in the output markets led each firm to charge price equal to marginal cost, but the second distortion would disappear.

It is not uncommon to hear the argument that market power cannot persist in these licensed markets so long as the market for the licenses is competitive. If there are 30 radio broadcasting licenses available in an area, how can there be much market power in the sale of advertising minutes? This argument confuses competition in the input market with competition in the downstream output market. The former does not have much if anything to do with the latter. If potential users of the input, in this case broadcasting licensees, have specific sunk assets that make certain uses of a license particularly attractive, *e.g.*, one applicant has an exhaustive collection of "new age" music, then they will be inclined to choose different uses of the license. The different uses will probably mean that they have little or no effect on one another in their respective output markets. Furthermore, even if there are no specific sunk assets, so that each licensee is equally well suited to serve any of the markets to which a license allows access, market power might still persist in the output markets. Consider, for example, a case in which the available frequency spectrum in an area can accommodate fewer radio stations than there are formats that could be profitably served (if the license were costless). In that case, some firms would probably operate as monopolists. A decision by any other licensee to "invade" a monopolized market will depend upon a comparison of the profits it earns in one of the duopoly or oligopoly markets with the profits it would earn as the second firm in the then-monopolized market. Clearly, the profits it earns under either decision are likely to be in part just transfers of income from the other firms in whichever market it chooses.

---

[4] Furthermore, these potential rents must persist for the incumbents after entry, so that given entry (or output expansion) of the new input holder and the resulting market price, the incumbents would still like to regain sales that they lost to the entrant. See Borenstein (1988).

[5] See Borenstein (1988) for a more complete examination of this dichotomy.

There is, however, one important distinction between the cases with specific sunk assets, and thus strong tendencies among license applicants to put them to certain different uses, and cases in which license applicants are more homogeneous. In the former case, the government can use the allocation of licenses among applicants as a tool to control final use of the licenses. In the latter case, control of final use can be done only through direct regulation of use; the regulators choice of applicant will not affect the use to which a license is put.

III. Pitfalls of Competitive Allocation of the Broadcast Spectrum

The previous section leads immediately to one question that is at the heart of analyzing market allocation of spectrum use: How closely are the profits of potential spectrum users correlated with the social value of their uses? Of course, a similar question must be asked of the obvious alternative approaches, *e.g.*, How closely are the license allocation choices of the government correlated with the social value of the uses to which each user will put its license? I will focus largely on the former question, but will state in advance that the latter must also be studied seriously before one could accept an administrative approach in lieu of market allocation.

The difficulties in analyzing the degree to which bids for use of spectrum space will reflect the social value of the intended uses can be separated into two areas. Some of the problems are generic to the sort of analyses necessary to estimate profit/social surplus correlations. Others are specific to the markets in which frequency spectrum is a critical input. Though both areas present formidable challenges, previous research has offered some guidance to dealing with the generic estimation problems. Problems specific to estimation of producer and consumer surplus in the broadcasting markets have received less attention and, at this point, loom larger.

Generic Problems in Measuring the Efficiency of Competitive License Markets

If the product that a broadcasting license permitted one to sell were a more traditional good or service, the analysis of market allocation of licenses would be theoretically straightforward, though empirically it would still be quite challenging. First, one would estimate the consumer surplus associated with the change in output that the license allowed a firm to effect in a given market. Second, an analysis of business stealing effects would be undertaken. If accurate measures of economic profit were available, this would be done most directly by measuring the change in the profits of incumbent firms in the market.[6]

If such profit data were not available, then one might approximate these sums by multiplying changes in the output of incumbent firms by their (dollar) markup. This will be a greater or poorer approximation depending on the elasticity of marginal cost with respect to output. If, as is generally the case, the marginal cost is less elastic in the short run than in the long run, then this approximation is likely to understate short-run business stealing effects by a greater amount than long-run business stealing effects.

In the generic case, the most severe difficulty probably lies in estimating consumer surplus in markets that a new licensee would serve as a monopolist. In that case, the consumer surplus attributable to the most inframarginal buyers – those with the highest reservation prices for the good – will play an important role in the estimation. This is not generally the case for most policy questions that one might attempt to address through demand estimation. In most demand estimation, one is interested in the shape and location of the demand curve around the equilibrium price and quantity. If total consumer surplus is to be estimated, however, then one must extrapolate from observed points to the extremely inframarginal price/quantity combinations that are seldom observed. In such cases, the extrapolation, and thus the estimates of total consumer surplus, is likely to be very sensitive to the functional form of demand that is assumed.

---

[6] Actually, the change in economic profits of all other firms who sell complements and substitutes should also be included. In very few cases, however, will the available data offer the depth or detail to allow extension of the empirical analysis to estimates of surplus changes in other markets.

In constrast, if the new licensee enters a homogeneous-good market in which incumbents are already present, then the change in output involves sales to people who have reservation prices below the pre-entry price and above the post-entry price. These bounds on the value of the product to the new purchasers allow more reliable estimation of the change in consumer surplus. A different, but related problem arises, however. If business stealing occurs, the cause may not be just the reallocation of sales in a symmetric Nash equilibrium, but rather the preference that some "incumbent buyers" have for the entrant's product over the incumbent's product. Such a preference is not easily accommodated by homogeneous-good theories of market interaction, *e.g.*, Cournot, or by the estimation procedures that rely on these theories. The increase in consumer surplus of buyers who switch brands may be nearly zero, but it may be quite large. In fact, as with the problem of inframarginal buyers in a monopoly market, there is no logical upper bound on the consumer surplus that the entrant offers these people.

One common approach to measuring the benefits of product differentiation has been to estimate hedonic prices for attributes of the goods in the market and then to infer the value of the new brand by calculating the difference between the sums of the values of its attributes and the price for which it is sold. Unfortunately, this approach nearly always involves estimation of common values (among consumers) for product attributes, primarily because the data available do not allow one to estimate distributions of these values across the population of consumers. Estimation of a single value (or value as a function of quantity of the attribute) fails to capture the value of the new entrant's product to those most inframarginal in their brand-switching decision. That is, there may be some buyers whose tastes lead them to strongly prefer the entrant's product over any of the incumbent's. If the average value of product attributes is used to calculate the entrant's contribution, while those who switch to the entrant value the attributes it offers more than the average person in the population, then the hedonic pricing approach will systematically undervalue the contribution of the new entrant.

Another approach to measuring the surplus gain of incumbent buyers due to entry is to estimate a general value of variety. Variety in product selection within a market may be valuable either because consumers have concave utility functions, as modeled by Dixit and Stiglitz (1977), and would therefore prefer to consume a range of products, or because they prefer a certain type of product in characteristic space and an increase in brands lowers the expected distance from their preferred point to the nearest brand. Salop (1979) has modeled such a market and Borenstein (1985) has generalized it to include consumers who are heterogeneous in their strength of brand preferences (*i.e.*, their costs of "traveling" a given distance in characteristic space). Though these models give a basis for arguing that variety has value to incumbent buyers and give structure to attempts to calculate that value, they still require some assumption or estimation of the parameters that determine the value of variety. As before, such parameters will be quite difficult to estimate, because they require information about demand that is not generally available from observing equilibrium outcomes.

Still, the prospects for empirical investigation of efficiency in license allocation are not as bleak as the previous paragraphs may make it seem. I would suggest two approaches that will mitigate the weaknesses that are clearly to some extent inherent in real-world welfare analysis. First, no market exists in complete isolation, and there are nearly always substitute products from which an upper bound on reservation prices may be approximated. An additional television station, for instance, is very unlikely to be worth more to incumbent viewers or inframarginal non-viewers than the alternative cost of expanding the station selection by subscribing to the local cable system.

Furthermore, one might be comfortable assuming that the value of an additional station is no greater for someone who switches to that station from another than for a consumer who enters the market only when the new brand becomes available. That is, the surplus gain to a new buyer might be assumed to be greater than the gain to the average incumbent buyer who then benefits from greater variety.[7]

Finally, even if precise estimates of some of the effects of license allocation are not available, estimates of certain other effects may give the policy maker a range of parameter values under which a proposed allocation will be preferred. For instance, assume that a license is to be used in one of two markets, depending on which of two applicants receives the license. Suppose that both markets have incumbent firms and that the effect of a new entrant on prices, total profits and total output in each market has been estimated. From these estimates, the analyst has been able to calculate the difference in profit changes and in the consumer surplus of new buyers between the two markets.[8] Reliable estimates of the surplus gain to incumbent buyers from increased variety are not, however, available. One could still come up with a comparison of the following hypothetical sort: The profit change will be $500 in market 1 and $1000 in market 2. The consumer surplus gain to new buyers in market 1 will be $1000 while it will be $2000 to new buyers in market 2. Thus, unless, the gains to incumbent buyers are $1500 greater in market 1 than in market 2, the license will increase total surplus more in market 2. The policy maker could then contemplate whether it is implausible that the omitted estimates differ by more than $1500, or whether this is still insufficient information to make a reliable decision.

In some cases, of course, little if any of the data necessary to carry out this analysis will be available or the decision maker may not trust the reliability of the estimates that analysts provide. In such cases, the policy maker might decide that the factors that determine the ratio of total surplus to the licensee's profits are not likely to differ substantially across potential licensees. As a result, he or she might choose to use the licensee's profits as the best indicator of relative social values of operation and, thus, choose to allocate the license through competitive bidding.

---

[7] Though this need not be true, it does seem reasonable and is consistent with the comparative analysis in a characteristic space model as presented in Borenstein (1986).

[8] The consumer surplus to new buyers might be approximated by assuming a linear demand curve, in which case $\Delta CS = 0.5(p_0 - p_1)(q_1 - q_0)$. Alternatively, the starting and ending prices and quantities may be used to bound the changes in consumer surplus, as demonstrated by Varian (1985).

Spectrum – Specific Problems of Efficiency in Market Allocation

Most of the previous section focussed on the general problem of license allocation under the assumption that the license allows the holder to sell an unlimited (or at least not strictly specified) quantity of a good or service directly to consumers. Licenses for use of the broadcast spectrum do not exactly fit this description. Some of the differences bear directly on the efficiency of competitive license allocation and on the analysis of efficiency. In this section, I discuss those differences and the importance they are likely to have in analyzing allocative efficiency with respect to the broadcast spectrum.

As pointed out previously by Steiner (1952), Spence and Owen (1977), and others, when producer revenues come from advertisers rather than from consumers directly, the relationship between profits and social surplus will change. Furthermore, advertiser financing of certain spectrum uses is in part a result of the public-good aspects of those media, another factor that increases the variance of the ratio of producer to total surplus in different uses of a license.[9] Spectrum rights also differ from operating licenses *per se*, because some services, *e.g.*, home television reception, can be provided even without use of the broadcast spectrum. Operating licenses *per se* are a necessary input for production of a good or service, while the broadcast spectrum may be just one input, without which the product can still be provided, albeit at a higher cost. Finally, the data for analysis of even the generic issues discussed in the previous section may pose particular problems in broadcasting markets.

---

[9] Advertiser financed operations are neither a necessary result of non-exclusivity, nor limited to such markets. Public television, financed in great part by voluntary donations, demonstrates that advertising is not necessary. The ads that are now often included at the beginning or end of rented home videos are evidence that such revenue sources are also tapped by distributors of private goods.

In order to talk specifically about uses of the spectrum that are non-exclusive, one must first define the output in such cases. For distribution of "free" television and radio, I will take as the output, the quantity of viewer time that the broadcast provides. This ignores the value to businesses and, possibly, to consumers, of the advertising content of the broadcast.[10] Though advertisements are less common in the "pay" uses of the media, *e.g.,* STV and cellular telephones, they do occasionally appear. I will also not consider the value of these ads to the advertisers and the consumers.

Ignoring differences in viewer demographics, advertisers that use broadcast media are interested only in the number of people that the ad reaches. Since the service that the producer is delivering to advertisers is viewer exposures, the producer is not interested in the enjoyment that a viewer receives from a show so long as he or she chooses to watch. Admittedly, it is also true in private markets that a producer is not concerned with consumer surplus, just profits, but greater per-consumer value of the product in private markets can be translated into higher profits (by raising price or more effectively price discriminating). In the "free" television and radio markets, the producer values greater viewer enjoyment only to the extent that it increases total viewership. Thus, as Spence and Owen pointed out, advertiser-financed media will tend to under-produce (relative to the social optimum) shows that are valued very highly by a narrow sector of the population. Competitive use of the spectrum range designated for over-the-air broadcasting is then likely to result in excess production of "mass appeal" programs.

This procedure for revenue collection in "free" television will not only cause greater variation in the profits to social surplus ratio across the possible programs in this media, it will also alter comparisons with this ratio in private uses of the spectrum. When the technology does allow privatization of the television programming product, *e.g.,* cable television, the profit-maximizing product is generally one with fewer ads and greater direct payments from the consumers than can be attained with over-the-air broadcasting. This supports the hypothesis that over-the-air broadcast stations capture a smaller portion of the total surplus that they generate than they would be able to if privatization were costless.[11] It does not necessarily follow immediately from the public good aspects of over-the-air broadcasting, or from the reliance on advertising, that these uses of the spectrum capture less of the surplus they create than do alternative uses of the spectrum that are private, *e.g.,* cellular telephone. Still, this does seem likely to be the case, and if it is, then private users will be willing to bid more relative to the social value of their output than producers of non-exclusive outputs. As a result, a market allocation of spectrum use across media is likely to lead to an excessive dedication to private use relative to non-exclusive use.

---

[10] Though this is consistent with FCC restrictions on ad minutes and with previous welfare analyses of the broadcast media, it is worth considering at some point the role that the values of ads themselves would play in the analysis.

Because broadcasters do not charge prices to their audience, it is not likely that one could reliably measure the consumer surplus. As explained in the previous section, however, one might still be able to make reasonable assumptions about the relative benefits to different audiences or to calculate the range of relative benefits that could justify a decision. One starting point for such an analysis would be to assume that incumbent radio or television audiences get no more additional consumer surplus from switching to a new station than do new consumers of the radio or television product. One might even give some smaller weight to incumbent listeners than to new audiences brought into the market. If this were the case, then a station that would generate an average audience of 10,000, most of whom would otherwise not listen to the medium, would be considered to make a greater consumer surplus contribution than a station that generated a similar size audience mostly from people who switch stations.

Not surprisingly, the same firms that offer value mostly to "brand-switchers" rather than new audiences are also the ones that generate a greater proportion of their profits from transfers from incumbent firms. Thus, a general bias towards "new" uses of the spectrum, particularly in assignment within a given medium, might well be welfare improving. Unfortunately, the optimal size of the bias is still, at best, a guess. Admittedly, this approach is disturbingly reminiscent of the arguments in airline route assignment under the CAB and other market assignments in other regulatory reviews in which the potential entrant was required to show that its operations would not just divert business from incumbent firms. There is an important distinction, however. Under the analytic approach suggested here, the harm to incumbents is not a considered a negative effect. Rather, the transfer of profits in such cases is simply not counted as a positive effect. To the extent that such transfers are accompanied by greater consumer satisfaction, or by additional profits due to greater production efficiency on the part of the entrant, these accompanying welfare gains clearly should be counted.

---

[11] The observation is also supported by the revenue mix that a similar private medium has settled upon. Newspapers generate more of their profits from ads than from subscription fees, but they do collect a substantial portion of their revenues from consumers. The proportion is still higher for magazines.

In one sense, the extreme example of audience stealing effects and negligible contribution to consumer surplus occurs when a spectrum right is used to provide a service that could be provided without the spectrum input, albeit at higher cost of production. Consider a potential provider of STV that plans to operate in lieu of a local cable company. Either provider would be a monopolist in the market, but the STV company would have slightly lower fixed costs. Assume, for simplicity that marginal costs would be the same for the two operations. Then the lowering of production costs by substituting spectrum for another input, *i.e.,* cable, will not reduce price to consumers. Thus, the social gains are not likely to be underestimated by the private valuation of spectrum in these uses, as there would be no gain in consumer surplus. At the same time, the profits from displacing a competitor's higher-cost operation might in large part be transfers, rather than new surplus. Thus, the probability is high that a bid to use spectrum in producing a service that could be produced without this input will overrepresent the social gains from this use of the resource.[12]

Finally, on the most gritty practical level of econometric analysis, some uses of the spectrum pose difficulties that are not present in analyses of most other license allocation issues. Not only might the data necessary for the analysis be competitively sensitive and therefore difficult to obtain from the firms involved, in the non-exclusive spectrum uses the producers of the product do not themselves know with much accuracy the number of consumers they are serving. Television data on viewership, for example, are subject to some noise, even when monitoring is done electronically and no log-keeping is necessary, because the monitoring device cannot detect the number of people (if any) who are watching the set.

---

[12] This is *not* an argument that there is small or no social value from using spectrum to produce a service that would still be created in the absence of the input. Rather, it refers to the relationship between the social value of such a use and the private value that would be represented in the competitive bid for the spectrum rights.

For a number of reasons, the error in measurement seems to be proportionally much greater with radio audiences. First, electronic monitoring of radio audiences is more costly than with television, because a typical family owns many more radios than televisions. Second, the private value of accurate measurement is smaller with radio, because ad rates (per thousand listeners, for instance) are much lower. There is simply less money at stake, but the survey or monitoring costs do not also decrease proportionally. Third, because people listen to the radio in many places, often while doing other things, they are less likely to keep accurate logs of their listening. Lastly, because most radio stations reach far fewer listeners in an area than any of the local television stations, the size of the survey necessary to obtain a similar standard error of audience size estimates *as a proportion* of the estimated size of the audience size will be much greater. A 0.5% margin of error may not be critical when a television station is watched by 15% of the local population, but it renders useless an estimate of audience for a small radio station that never obtains more than a 1% share.

IV: Conclusions

With all of the difficulties in measuring the efficiency consequences of allocating spectrum rights in a competitive market, is there likely to be any payoff in attempting such studies? My answer is a very qualified yes. There are certain instances in which a thorough analysis of the relationship between private and social value of different spectrum uses would be enlightening.

More importantly, however, is the recognition that in any process of market allocation of spectrum, there can be a significant variance across uses in the ratio of private to social value of the use. Furthermore, there are characteristics of the uses that make it particularly likely that a bidder will be willing to offer a greater or smaller proportion of the change in the total surplus that his or her use will create. Recognition of these characteristics make it possible for the policy maker to perhaps avoid some of the most adverse outcomes that could result from market allocation.

In these times when government resources grow scarcer by the day, one must ask whether the benefits from such investigations are likely to outweigh the costs. Unfortunately, we will not know the answer to this question until some resources are expended.[13] Some guidance in direction of those resources can be gotten from the applied theory.

First, when the technology for a certain use of the spectrum makes its product non-exclusive, there is a strong case to be made that the firms will capture an unusually low proportion of the new social surplus that they generate. Further work in quantifying the benefits of non-exclusive use versus exclusive use of the spectrum, and the private to social surplus ratios, should be a high priority. Clearly, this is not an issue if the allocation of spectrum across media is fixed, but the greatest social gains from market mechanisms or other programs for reallocation may come from exactly such reallocations across media.

Second, when a firm's operations generate few new users of the product, *e.g.*, a seventh top-40 radio station, but instead result primarily in brand switching, there is reason to suspect that its private returns from that use of the spectrum may be far in excess of the social gains. Here, further work is also in order, but the outlook for strong findings is less sanguine. At some point administrators may have to make a decision regarding the value of variety within a given format or type of spectrum use. One decision might be to follow the guidance of a competitive market for spectrum. Given the costs and limitations of the research necessary to measure social benefits, this may be the preferred approach. Still, such a step should be taken with the understanding that it is not a panacea, but a considered choice among a spectrum of imperfect alternatives.

---

[13] Some enthusiasts of market allocation in all cases might argue that the last 50 years of FCC attempts to guide spectrum use has shown us that the costs are substantial and the benefits are negligible if they are even positive. As theoretical and empirical techniques become more refined, and as more data become available, one would think that such guidance could be done more sensibly. If that is not the case, supporters of market allocation of all spectrum have a strong case.

REFERENCES

Borenstein, S., "Price Discrimination in Free-Entry Markets," *Rand Journal of Economics*, **16**(Autumn, 1985).

——————, "On the Efficiency of Competitive Markets for Operating Licenses," *Quarterly Journal of Economics*, **103**(May, 1988).

——————, "Too Much Rock and Roll?: Business Stealing Effects in Radio Station Format Choice," presented at the 1986 Winter meetings of the Econometric Society (December 1986).

Dixit, A.K. and Stiglitz J.E., "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, **67**(June 1977).

Geller, H., "The Role of Future Regulation: Licensing, Spectrum Allocation, Content, Access, Common Carrier, and Rates," in Noam, ed., *Video Media Competition: Regulation, Economics, and Technology*, New York: Columbia University Press, 1985.

Kwerel, E. and Felker, A.D., "Using Auctions to Select FCC Licensees," Office of Plans and Policy Working Paper #16, Federal Communications Commission, (May 1985).

Spence, M., "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, **43**(June 1976).

——————, and Owen, B., "Television Programming, Monopolistic Competition, and Welfare," *Quarterly Journal of Economics*, **91**(February 1977).

Steiner, P., "Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting," *Quarterly Journal of Economics*, **66**(May 1952).

Salop, S., "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, **10**(Spring 1979).

Varian, H.R., "Price Discrimination and Social Welfare," *American Economic Review*, **75**(September 1985).

Divestiture Policy Considerations in an Information Services World

1988 Telecommunications Policy Research Conference

Airlie House — October 31, 1988

Timothy J. Brennan

Associate Professor, Public Policy, Communication and Economics

George Washington University.

INTRODUCTION

An important component of the settlement of the Justice Department's antitrust case against AT&T was that the divested local Bell operating companies (BOCs) were prohibited from providing "information services," These are defined as "the offering of a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing, or making available information which may be conveyed via telecommunications...."[1] Less than two years after the divestiture in January, 1984, however, the Justice Department reversed itself and advocated removal of this prohibition. In the FCC's "Computer III" proceeding, Justice advocated that a BOC be permitted to provide "ancillary services" [including information services] if it provides "others the same type of interconnection to the [local telephone] bottleneck that it provides itself," unless the BOC can show "significant efficiency losses" or that the ancillary service is a "natural monopoly."[2] The Department subsequently advised the divestiture court that the restriction on BOC provision of information services be rescinded, arguing authority over the "difficult line-drawing issues" should be left to the FCC.[3] The FCC has advocated that the BOCs (and AT&T) be permitted to offer information ("enhanced") services without formal structural restrictions, relying upon accounting safeguards and equivalent interconnection policies to prevent abuses.[4]

---

[1] United States v. American Telephone and Telegraph Co., 552 F. Supp. 131, 227, 229 (D.D.C. 1982) ["Divestiture Decision"].

[2] U. S. Department of Justice, Comments Before the Federal Communications in the Matter of Amendment of Sections 64.702 of the Commission's Rules and Regulations (Third Computer Inquiry); CC Docket No. 85-229, Nov. 13, 1985.

Despite the widely held view that the BOCs can provide information services without realizing the anticompetitive potential that motivated the Justice Department to recommend the restrictions in the first place, the court has elected to continue to keep the BOCs from providing information *content*.[5] With respect to information *transmission*, the court has concluded that certain "infrastructure components" necessary for the communication of other firms' information may have been blocked by the original MFJ restrictions.[6] After soliciting additional submissions, the court concluded that the BOCs should be allowed to provide "data transmission, address translation, protocol conversion, billing management, and introductory information content," as well as voice storage and retrieval and electronic mail.[7]

---

[3] U. S. Department of Justice, Report and Recommendations of the United States Concerning the Line of Business Restrictions Imposed on the Bell Operating Companies By the Modification of Final Judgment, Feb. 2, 1987 ["DOJ Triennial Report"]: 134-51.

[4] Federal Communications Commission, Amendment of Sections 64.702 of the Commission's Rules and Regulations *(Third Computer Inquiry) Report and Order*, 104 FCC 2d 958 (1986).

[5] U.S. v. Western Electric Company, Inc., et al., Order, Civil Action No. 82-0192, *slip op.* at 115; Sept. 10, 1987 ["Triennial Review Order"].

[6] *Id.* at 188.

It appears the BOCs are going to be permitted to enter at least some information services markets expressly forbidden to them just five years ago, with wider entry forecast for the future. As technology changes, the information handling capacity of local telephone networks may increase dramatically, permitting simultaneous transmission of voice, video, and high-speed data communications.[8] The FCC has recently proposed permitting the telephone companies to offer cable television service, not merely as a video transmitter but as a packager of program content.

These developments do not mean the considerations originally underlying the information services restriction no longer apply. The efforts of the FCC in the Computer III proceeding, for example, reflect the view that some regulatory oversight of BOC activities related to information services remains necessary. The purpose of this paper is to review a variety of policies that might be relied upon to address these considerations as the BOCs are allowed to integrate into information services. After reviewing the purpose of keeping the BOCs out of those markets, we look at a variety of non-structural policies that may be used to prevent anticompetitive harms. These include *ex post* antitrust enforcement, cost-based rules, regulatory reforms such as accounting procedures and "Open Network Architecture" policies, and predatory pricing tests. We end with conjectures as to why information services may not have blossomed to the extent some have foreseen, in particular, whether current structural regulations are to blame. Our goal here is not to reach a conclusion on the optimal policy toward BOC information services provision, but only to improve understanding of the more pertinent issues through a survey of the wide range of possibilities.

---

[7] U.S. v. Western Electric Company, Inc., et al., Order, Civil Action No. 82-0192, *slip op.* at 65-66; Mar. 7, 1988 ["Information Services Order"].

[8] Heidman, R. K., *Telecommunications Management Planning ISDN Networks Products and Services* (Blue Ridge Summit, PA: Tab Books, 1987): 264-286.

WHY KEEP BOCS OUT OF INFORMATION SERVICES MARKETS, AND WHY NOT?

The problems associated with BOC information services offerings follow from the general theory of how a regulated firm with market power may exploit its position through affiliated enterprises in other unregulated markets. This theory has been discussed at length elsewhere;[9] only its highlights will be summarized here.

A. *Cross-Subsidization*

Under conventional regulation, a regulated firm's prices are based upon the costs the regulator determines were incurred to provide the regulated service. If these costs rise, so do the prices. A regulated firm therefore has the incentive to let the regulator believe costs are higher than they actually are. Normally, fabricated cost data would not be submittable without being able to show that the cost data reflected actual purchases of inputs. If the inputs were used to produce the regulated service, though, there would be no profit in purchasing them. If the regulator permits higher prices, the added revenue would merely cover the added costs of the purchased inputs, with no gain to the firm.

---

[9] Brennan, T., "Why Regulated Firms Should Be Kept Out of Unregulated Markets: Understanding the Divestiture in *U.S. v. AT&T*," *Antitrust Bulletin* 32 (1987) : 741-93.

On the other hand, suppose the purchased inputs could be used in some other business. If the regulator continues to believe that the inputs are being used to provide the regulated product, it will permit the regulated firm's revenues to increase to cover their costs. The inputs, though, can be used to produce other goods or services. As usually phrased, the production of the other goods or services is now being "cross-subsidized" by the ratepayers, in that the regulator is having the ratepayers pay for these inputs. The firm gets to raise price in the regulated market; the profits show up as a reduction in the costs of serving other markets. Note that "cross-subsidization" is the *end* and serving other markets is the *means* to that end; in contrast to the common perception that cross-subsidization is but a tactic for the monopolization of an unregulated market.

The effects of cross-subsidization are that the regulated market's price rises and the regulated firm's outside operations become inefficiently large. In addition, as a result of the subsidized reduction in the marginal cost of serving other markets, the cross-subsidy may create a credible threat to exclude or predate against more efficient firms. For cross-subsidization to be profitable, though, three things are necessary. First, the regulatory link between reported costs and prices must be fairly tight, to ensure that regulated rates rise as these costs are reported to the regulator. Second, the regulated firm must have "market power" in its regulated market. Otherwise, competition will keep it from profitably raising prices to the level the regulator would allow in order to cover "costs". Third, prices in the other goods and services markets in which the inputs are really used must not be regulated closely. If the apparent cost of producing these goods decreases, because the inputs are charged to the regulated accounts, the regulator could cause price to fall in these "other" markets. Such a price decrease would take away the revenue gained by causing price to rise in the cross-subsidizing market, eliminating the profit motive for the cost misallocation.

A variety of tactics might be used by regulated local exchange companies to cross-subsidize operations in information services markets.[10] The basic strategy entails placing costs of information services operations into the regulated account. In terms of content generation, Peter Huber writes:

> [T]he preparation of the information content itself is one activity that LECs [local exchange carriers] would have notably little opportunity to cross-subsidize. Running a local telephone exchange does not require reporters, copy editors, joke writers, financial analysts, astrologists, or other information content providers.[11]

Huber's implication is that a BOC could not place costs of generating content onto the regulated account, because the regulator could not be fooled into thinking costs of these inputs could be used to provide local telephone service. While this may not be a completely accurate inference—I assume BOCs use financial analysts and editors[12]—the point is well taken as far as it goes. However, there are other inputs into the generation of information content that may be used in common with the provision of local exchange service. One such class of inputs is technical equipment, such as word processors and file servers. A second and important class of inputs is financial capital. While the BOC could not simply borrow money ostensibly for investment in telephone service and use it to seed information generation businesses, it could shift some of the borrowing costs to the ratepayers to the extent telephone assets or earnings are used as collateral for the investments in other industries.[13]

---

[10] A formal discussion of the general incentives regulated firms have to engage in these tactics may be found in Brennan T., "Cross-Subsidization and Discrimination by Regulated Monopolists," *EAG Discussion Paper 87-2*, Antitrust Division, U.S. Department of Justice (1987).

[11] Huber, P. *The Geodesic Network: 1987 Report on Competition in the Telephone Industry* (Washington: U.S. Department of Justice, Antitrust Division, 1987) ["Huber Report"]: 6.35.

[12] Huber jokes that they may use "astrologists" [sic] for "gauging the regulatory and legal winds." *Id.*

Cross-subsidy becomes more complex when we turn to the technological side of the information services industry. Costs of inputs used to provide electronic data processing might be easily allocated to the regulated telephone business since, as the aphorism regarding the merger of communications and computing indicates, the technologies used to provide either service are similar. For example, computers and electronic equipment used to provide an information service could be used to assist in the provision of exchange telecommunications as well.[14] Other non-technological inputs into local exchange service, such as location, marketing, and billing, as well as research and development expenses, could be improperly allocated to the local exchange ratepayers.[15]

---

[13] Suppose, for example, that the cost of capital for local telephone service is 10% but, because of risks associated with competition and demand uncertainty, the cost of capital to enter an information service business is 15%. If loans to enter the information business are backed by the nearly certain earning capacity of the telephone business, the firm could borrow at essentially the cost it takes to raise capital to provide telephone service. However, since part of the telephone investment now is dependent upon riskier information service revenues, the cost of capital for this LEC rises above 10%.

The size of the increase depends upon various statistical factors, but its effect under cost-of-service regulation is to increase local phone rates. The "cost" of capital for the information service business, though, is reduced, as far as the telephone company is concerned. In effect, the difference between 15% and the lower cost of capital is covered by the payments from the ratepayers, following the increased cost of capital for telephone operations.

[14] Triennial Review Order at 112.

B. Discrimination

The problem more on the minds of courts and other policy makers springs from a second strategy to evade price regulation. A firm forced to hold its prices down below the monopoly level has an incentive to deny the regulated product to those who do not buy another product from it at a premium above the general market price. If tying availability of the regulated product to the purchase of an unregulated product can be brought about, the firm can capture the difference between the regulated price and the market price for its regulated product through the premium charged for the other product. A favorite example of this is where, in cities with anti-scalping laws, someone sells Bruce Springsteen tickets at face value, but only if one buys his "rare" empty Budweiser can for $100.

Such a transparent evasion of price control would probably not avoid prosecution; neither would an attempt by an LEC to deny service to those who, say, didn't rent shoelaces from the LEC for $25 a month. In the LEC's case, though, there may be more subtle methods to make local exchange service contingent upon purchasing LEC products in other markets. The first step is to look toward other products which require local exchange service to function. Three such products mentioned in the Modified Final Judgment in

---

[15] This list is derived from the Huber Report at 6.36. Note, though, that the discussion in the Huber Report discusses not cost *misallocation*, but actual cost *savings* that may be realized if a telephone company enters an information services market. If these cost savings can be achieved only through corporate integration, they are an important potential reason to let BOCs into information service markets, not to exclude them.

U.S. v. AT&T ("MFJ") are customer premises equipment, interexchange services, and information services.[16] Simple denial of LEC service to others selling such products or purchasing them from non-LEC vendors would presumably be no more feasible than the shoelace rental scheme. However, the LEC could make life difficult for other suppliers or their customers, because of the technical cooperation needed for other entrants to provide their product successfully.

Such technical cooperation is of paramount importance to those information service vendors who require connection to local telephone exchanges to reach their customers. Among the concerns expressed have been favoritism in local network design, access to proprietary information regarding network design or customer needs, and provision and maintenance of access lines.[17] By its nature, data transmission is especially susceptible to error, making proper engineering cooperation a paramount concern for entrants into information services markets. While there exist other means for distributing information, broadly defined, such as satellites, cable and over-the-air television, mobile radio, and print media,[18] two-way or switched transmission, especially for those without the volume warranting installation of private lines or transmission facilities, require LEC access.[19]

Discrimination against non-LEC information service providers ("ISPs") in the provision of LEC access would have harms similar to those associated with cross-subsidization. The prime effect of discrimination would be higher prices in for those information services which require LEC access for efficient provision, thus reducing the competitive performance of those markets. A secondary effect might be the limitation of operations or exclusion of unaffiliated ISPs more efficient than the BOCs. A "victim" of discrimination might have been able to provide a more desired information service at lower cost; its exclusion further inhibits efficient performance of the information services industry.

---

[16] Divestiture Decision at 227, 231.

[17] Triennial Review Order at 103, 112.

[18] Huber Report at 6.17-22.

[19] Triennial Review Order at 107-108.

C. The "Quarantine" Solution and Its Problems

The MFJ solution to the problems of cross-subsidization and discrimination affecting information services markets was radical—keep the BOCs out of those markets. If the BOCs are not in information services markets, they are not in a position to profit from charging inputs used elsewhere to the accounts for its regulated services. There is no mechanism to raise regulated rates by misallocating costs, while taking the profits through the "subsidy" of operations in other markets. Similarly, excluding the BOCs from information services markets eliminates the incentive to discriminate. If they are not in such markets, they have no incentive to make life miserable for unaffiliated ISPs; there is no affiliated ISP to profit from the woes of competitors.[20]

As noted, the trend is to look to less radical solutions. The primary problem with excluding the BOCs from information services markets is that they may well be the most efficient providers of some of these services.

---

[20] Triennial Review Order at 103; U.S. Department of Justice, Competitive Impact Statement in Connection With Proposed Modification of Final Judgment, 47 F. R. 7170, 7179-80 (Feb. 17, 1982); Brennan *supra* n. 9.

There may be "economies of scope" between local exchange service and an information service; i.e., it may be less costly for the same firm to produce both local exchange service and the information service than to provide them separately. If so, keeping out the BOCs would be a wasteful refusal to take advantage of cost savings attainable where there are economies of scope. For example, the possibility that providing protocol conversion or billing information is less costly for BOCs than for consumers or independent providers would be a strong justification to allow BOCs to provide the service. In the extreme, the information services restriction may keep an information service off the market entirely if the BOC is the only cost-effective provider. If the information service is currently provided by only one or a few firms, the increase in competition in the information service market may provide economic benefits exceeding the expected harms to ratepayers or market performance from discrimination or cross-subsidization of that market.[21]

A good illustration of this may be cable television. Under current policy, a BOC is permitted to construct cable facilities for the use of independent providers, but is not allowed to provide the programming or market the cable *per se*. FCC rules have excluded telephone companies from cable service, on the grounds that a telephone company would discriminate against unaffiliated cable providers in access to facilities needed to provide cable, particularly attachments to telephone poles or space in underground wire conduits. An exception was made in low density rural areas where, supposedly, only the telephone company would provide cable service.[22] The "telco-cable cross-ownership" ban and the rural exception were codified in the 1984 Cable Act. Recently, however, the National Telecommunications and Information Administration proposed amending the FCC rules to allow the BOCs to offer a "video dial tone" in which a customer could access television through the local telephone network. The FCC has gone further, proposing permitting telephone entry into cable. Envisioned changes in local exchange technology such as ISDN, which would permit efficient provision of video through the telephone system, may be increasing economies of scope between voice and video transmission. Video transmission might be made more competitive as well, since cable is almost ubiquitously provided by only one firm per geographic market.

---

[21] Conversely, if the information service is provided efficiently and competitively, there is little need to invite cross-subsidization and discrimination by permitting a telephone company to provide the service.

Similarly, while there may be little harm in permitting a BOC to provide an information service that does not rely on local exchange facilities for transmission or does not use inputs similar to those used for local service, the likelihood of substantial economies of scope in such cases is also not great. The principle of minimal government interference suggests that the BOCs should not be prohibited from offering such services, but the costs or benefits on either side of the ledger would usually not be significant.

If the policy trend is to put more emphasis on economies of scope and less on the possibility of anticompetitive conduct, then BOC entry into information service markets can be expected to become greater over time. It may thus be useful to look at alternative policies to protect against discrimination and cross-subsidization.

---

[22] The rural exception policy presumes there are economies of scope between telephone and cable service are significant at small levels of operation but not at large levels.

ALTERNATE POLICIES TOWARD INTEGRATED REGULATED FIRMS

A. *Ex Post Antitrust Enforcement*

1. *Antitrust as Deterrent*

Given the history of telephone regulation, a surprising policy response to the potential for cross-subsidization and discrimination might be to use conventional "after-the-fact" antitrust enforcement as a deterrent.[23] Rather than attempt to structure regulation to prevent antitcompetitive conduct *ex ante*, we could simply prosecute those LECs found to have been misallocating costs of information services to the regulated ratepayers or improperly favoring affiliated information services providers over their competitors in access to exchange telecommunications.

In thinking about a law enforcement approach, a story comes to mind that an attorney and former colleague once related, regarding rich "Uncle Sidney." Suppose you have a rich Uncle Sidney, who has mentioned you favorably in his will. Suppose further that you go out and buy a rifle. You then would have both the incentive to murder Uncle Sidney and the ability to do so. If murder were treated in the same way as antitrust, you would be subject to legal stricture as soon as you bought the rifle, having acquired the ability to act on the incentive. You might be subject to structural regulation, requiring that the rifle be used only by a fully separated subsidiary, or, in the spirit of the MFJ, no one with a rich uncle could be permitted to own a gun.

---

[23] "Antitrust enforcement" includes both Justice Department and Federal Trade Commission action under the antitrust laws, as well as FCC actions taken in response to improper LEC conduct rather than as before-the-fact regulations.

There are a host of differences between murder and antitrust, not the least of which is that cross-subsidization isn't a capital crime. The interesting distinction here, though, is that most of us would not automatically assume that an armed nephew would necessarily murder his uncle, while economists seem to assume that a firm in a position to act anticompetitively will do so. Why the difference? The main one seems to that we find it sensible to rely in part on conscience as a deterrent, a notion ludicrous to most economists at least in antitrust contexts. This may be an appropriate distinction. Excepting gross collusion, the boundary lines between "good" and "bad" conduct in antitrust, when the issue is one of efficiency rather than, say, life or death, can be too fine to warrant moral indignation. However, if antitrust isn't taken seriously, a firm with the opportunity to cross-subsidize or discriminate may be expected to try it; conscience is not likely to get in the way. After all, if conscience would work, would regulation be necessary in the first place?[24]

Another threshold issue is that the harms addressed by the MFJ may not be addressable through antitrust enforcement. It is important to remember that *U.S. v. AT&T* was settled, not decided; the record on the private cases against AT&T is mixed. Moreover, the Department of Justice's current view is that harm caused to ratepayers by cross-subsidization is not a matter for antitrust prosecution.[25] While this view is not universally held, it adds to the doubt that antitrust enforcement can necessarily take care of the problem.

---

[24] It may be interesting to note the view that the pre-divestiture AT&T was motivated by public service, not profit. See for example Temin, P., *The Fall of the Bell System* (New York: Cambridge University Press, 1987): 3.

Another interesting distinction between antitrust and "Uncle Sidney" is that detection that a murder has been committed is usually easier than detecting whether a welfare-reducing antitrust violation has occurred. I owe this observation to Mary Fitzpatrick.

2. Are Cross-Subsidization or Discrimination Still Attractive Options?

A further issue is whether LECs will find cross-subsidization of discrimination strategies attractive. Faulhaber provides the strongest criticism of the cross-subsidization story.[26] The cross-subsidy story requires a strong combination of regulatory diligence and ineptitude. The "diligence" is that rates must be strictly tied to costs actually incurred; the "ineptitude" is the inability to determine which costs were actually incurred in the provision of the regulated service. Faulhaber focuses on the apparent contradiction between diligence and ineptitude, arguing that a regulated firm able to fool the regulator regarding cost allocations probably is so loosely regulated that it need not go through this elaborate accounting ruse to exercise market power; it can do so already. A regulator insufficiently competent to determine which inputs are used to produced which outputs can probably be lied to directly; the regulated firm could just claim high costs without having to actually purchase inputs. An inept regulator probably isn't tying rates too closely to actual costs in the first place.

---

[25] DOJ Triennial Report at 47-49.

[26] Faulhaber, G., *Telecommunications in Turmoil* (Cambridge, MA: Ballinger, 1987): 118-20. His criticism is all the stronger because, unlike many critics, he seems to understand that cross-subsidization is a tactic for evading regulatory constraint through cost misallocation, not either a form of predation or an "unfair" allocation of the costs of facilities used jointly to produce regulated and unregulated products.

Faulhaber's argument makes it clear that the combination of diligence and ineptitude is needed for cross-subsidization to be a matter of concern. Whether concern is reasonable is very much in the eye of the analyst. In the best "test case," *U.S. v. AT&T*, the Justice Department essentially never was able to prove that cross-subsidy occurred. Part of the difficulty may have been that there had been a social policy to subsidize monopoly local exchange operations with revenues raised from the increasingly competitive long distance market,[27] but part of the problem may have been the sheer difficulty in obtaining and interpreting AT&T cost data. Nevertheless, it remains at least plausible that the more cost data the firm can report, and the more the data correspond to actually incurred expenses, the higher would be rates under cost-of-service regulation. Moreover, the Justice Department successfully persuaded the court that AT&T had charged prices in other markets with such impunity that it must have been a corporate expectation that the costs would eventually be covered—"pricing without regard to cost"—revealing intent to monopolize.[28]

Thus, there is some evidence, albeit weak, that a firm holding a monopoly position in exchange telecommunications will succumb to the temptation to cross-subsidize. The inability to generate confirming data, and the need to rely on innovative theories of monopoly intent rather than conduct, suggests that *ex post* antitrust enforcement may not effectively deter cross-subsidization. The Justice Department's current view that the only harm relevant for antitrust purposes is predation in the unregulated market, not higher rates in the regulated market, only adds to the skepticism.

---

[27] Temin *supra* n. 24 at 357-59.

[28] U. S. v. AT&T, Opinion on AT&T's Motion To Dismiss, ("Opinion on Motion To Dismiss") *slip op.* (1981) at 43-53.

Like cross-subsidization, discrimination requires a paradox to be a profitable tactic. On the one hand, discrimination works only if the regulator "permits" it. Assuming the regulator is disposed to prevent evasion of price regulation through explicit discriminatory access, this "permission" could be secured only implicitly by deception, implying the regulator cannot detect the discrimination. On the other hand, discrimination pays only if it serves to boost demand for the regulated firm's affiliated ISP, at the expense of unaffiliated competitors. Such a demand shift cannot occur unless discrimination leads consumers to perceive that the affiliated competitor provides a more valuable product, or the unaffiliated competitors provide an inferior or more expensive product. In short, discrimination works only if the market perceives it while the regulator does not.

For example, suppose the LEC elects to maintain access lines to its affiliated ISP at a higher signal-to-noise ratio than those to competing ISPs. This would pay only if the quality difference were translated into higher prices consumers were willing to pay for the affiliated ISP. However, if the quality differences are sufficiently systematic to get consumers to perceive a difference between affiliated and unaffiliated ISPs, the regulator ought to detect the difference as well and, one might think, take enforcement action.

U.S. v. AT&T is instructive on this account. Most of the episodes of discrimination cited in the case, such as cutting off of MCI lines,[29] were not at all subtle.[30] They could only have been done in an era in which such regulation could have seemed tolerable to a regulator. Much of the case was not about whether AT&T "did it," but whether it was justified in doing so by explicit FCC regulation or implicit FCC policy regarding entry and the public interest in maintaining a monopoly network. As I believe Richard Levine once said, paraphrasing Eric Segal, "Regulation means never having to say you're sorry." While the trial judge in *U.S. v. AT&T* seemed not to view regulation as an excuse, other judges did not always see it the same way.

The trial experience suggests the discriminatory conduct of the past was conducted in an era when the regulator would be thought of as a cooperative partner. If we are not in such an era now, discrimination may be unlikely because of it is caught between the rock of regulatory vigilance and the hard place of market perception. If discrimination is unlikely, no policies other than *ex post* enforcement may be needed.

But the situation may not be so simple. Neither antitrust enforcement nor regulatory action are timely processes. Discrimination may persist for so long that by the time legal recourse is achieved, the victims are out of business. More importantly, the complexity of telephone technology and the LECs control over information may make it difficult for the regulator to attribute performance differences between affiliated and unaffiliated ISPs to improper discrimination. The LEC may claim that the performance differences are a result of economies of scope or the happenstance that its ISPs have the best engineering. The LEC profits as long as consumers perceive a quality difference; the consumers do not have to perceive the discriminatory conduct itself. If the issue can be so confused, there is some probability that enforcement will let some discriminatory conduct go unpunished, and a greater probability that *ex post* enforcement proceedings will be protracted. Finally, if the "capture theory" of regulation is to be believed,[31] there is no assurance that over the long term communications industry regulators can be trusted to police against discrimination.

---

[29] Opinion on Motion to Dismiss at 34-43.

[30] Depending upon how one interprets "discrimination," an important exception was the allegation that AT&T improperly favored Western Electric in its purchases of telephone equipment. This tactic is more like "cross-subsidization," in that it involves fabrication of artificially high transfer prices, to boost regulated rates and shift profits to an unregulated affiliated upstream input supplier. Brennan *supra* n. 10.

To summarize, there are good reasons to believe that concerns regarding cross-subsidization and discrimination motivating *ex ante* structural solutions may be exaggerated. Both require an apparently unlikely combination of circumstances to be profitable. Cross-subsidization requires diligent but incompetent regulators; discrimination requires perceptive consumers but unperceptive regulators. *Ex post* enforcement proceedings, however, may be too protracted or error-prone to be effective deterrents against whatever remaining risk of harm is presented by BOC entry into information services markets.

---

[31] Stigler, G., "The Theory of Economic Regulation," in Stigler, G., *The Citizen and the State* (Chicago: University of Chicago Press, 1975).

B. Regulatory Reform

1. Redesign

Since cross-subsidy and discrimination are the byproducts of regulation, reforms of telephone regulation may serve to lessen their likelihood, reducing the need for prohibiting or conditioning BOC entry into information service markets.

a. Price Caps

Recall that cross-subsidization requires that rates be closely tied to reported costs. Otherwise, there is no incentive for the firm to misallocate cost to the regulated business; doing so would not cause a reduction in rates. Recently, the FCC has proposed replacing cost-of-service regulation, where rates are tied to costs, to a form of regulation where the regulated firm can charge up to a "price cap" regardless of what it's costs actually are.[32] If "price cap" regulation were to go into effect, the risk of cross-subsidization would be virtually eliminated since, by definition, costs have no effect on price.[33]

---

[32] The FCC (Notice of Proposed Rulemaking in the Matter of Policy and Rules Concerning Rates for Dominant Carriers, CC Docket No. 87-313 , 1987). The use of price caps to regulate AT&T post-divestiture was suggested by the Department of Justice, Comments Before the FCC In the Matter of Long-Run Regulation of AT&T's Basic Domestic Interstate Services, CC Docket No. 83-1147 (1984).

[33] Haring, J. and Kwerel, E., "Competition Policy in the Post-Equal Access Market," *2 FCC Rcd* 5 (1987): 1488-93; Brennan, T. "Capping 'Average' Prices of Regulated Monopolists," Department of Economics, George Washington University, 1987.

Unfortunately, "price cap" regulation's very virtue, the separation of price from costs, implies that over time there is some likelihood that the regulated firm's profits may become either exorbitant or inadequate by conventional economic standards. For this reason, it is politically improbable, if not legally impossible, for "price cap" regulation to be implemented for a firm that will be regulated for some time to come.[34] Somewhere down the line, profits will influence policy, with rate reduction if profits are excessive or rate increases if profits have not been "just and reasonable."[35] To the extent LEC profits affect rates, and incurred costs affect the data telephone regulators use to calculate profits, the incentive to cross-subsidize information services is reestablished.

Even if "perfect" price cap regulation were instituted, a regulated firm would retain an incentive to discriminate. Discrimination does not require that rates be tied to reported costs, only that they be held below the profit-maximizing level.[36] As long as the exchange telecommunications market would profitably bear a higher price, a regulated LEC will have an incentive to get people to pay more for its services through paying excessive prices for complementary products such as information services.

---

[34] Brennan, *supra* n. 33.

[35] The FCC has elected to review price caps on the basis of profits. FCC, Further Notice of Proposed Rulemaking in the Matter of Policy and Rules Concerning Rates for Dominant Carriers, CC Docket No. 87-313, (1988): ¶128.

[36] Brennan, *supra* n. 10.

b. Open Network Architecture

The remedy for discrimination proposed by the FCC is "Open Network Architecture" ("ONA").[37] The purpose of ONA is to prohibit discrimination by requiring LECs design their networks so that any ISP can get pick and choose among a set of minimally defined "service elements" just as an LEC's affiliated ISP might, with no difference in function, performance, or cost. The point is to guarantee "equal access" to ISPs, just as interexchange carriers have equal access under the MFJ. With standardized interfaces and prices, the ability of an LEC to discriminate in favor of an affiliated ISP would be limited.

Whether ONA can meet these needs remains to be seen. In the context of divestiture, ONA puts the technological cart before the economic horse. Recall that reason to eliminate restrictions is to realize economies of scope available to the BOCs. However, ONA, with the requirement that all ISPs be treated equally, implies that the local exchange network will be designed to prevent most if not all economies of scope between local exchange service and information service provision from being achieved. But if operational economies of scope are nonexistent, there may be little if any reason to risk discrimination or cross-subsidization by letting an LEC into the information service markets addressed by ONA.[38] The best case for letting LECs into an information service market is where ONA is impossible, i.e., the LEC is for technical reasons the best if not only provider of that particular service. One may contrast ONA with equal access policy for interexchange service, where the BOCs were ordered to provide equal access without being invited to enter that market.

---

[37] FCC, Amendment of Section 64.702 of the Commission's Rules and Regulations (Third Computer Inquiry), CC Docket 85-229 (June 16, 1986).

Be that as it may, will ONA counter discrimination? Since it has not yet been implemented, we are left with speculations.[39] Clearly the view that ONA will work is not unanimous; criticisms of ONA so far include improper bundling of local exchange services, refusal of the BOCs to agree on standard service definitions, and a debate over whether the BOCs should be obligated to place other ISPs' equipment in the central office, if not within the switch itself.[40] Note further that ONA obligates the BOCs to offer only what they use themselves. An ISP desiring a differentiated form of exchange access need not be served, and an LEC providing a competing information service would have an incentive to deny or delay provision of this sort of access.

---

[38] Nonoperational areas involving marketing or research and development may generate economies of scope. Marketing economies may be thought of as gains from "one-stop shopping," but these gains could be achieved without LEC entry into information services markets if ISPs or other communications firms were permitted to resell local exchange service. R&D economies of scope are trickier. To the extent LECs make discoveries regarding information services in the process of doing R&D for regulated exchange services, the knowledge gained could in principle be licensed to independent providers. Since such R&D is often ratepayer-funded, the proceeds could be used to reduce local service costs.

[39] Triennial Review Order at 139-42.

[40] Hatfield Associates, *Open Network Architecture: A Promise Not Realized* (Boulder: Hatfield Associates, 1988),

2. *Cost-Based Rules*

a. *New Developments in Accounting*

Since cross-subsidization involves cost misallocation, the direct method of prevention would be to check to see that costs of doing business in information services markets are not improperly allocated to regulated local exchange accounts. Since earlier cost-based regulation of AT&T's offerings in response to competition were notoriously slow and unsuccessful,[41] one would need to believe circumstances had changed to believe accounting controls would be effective now. Changed circumstances can fall into one of two areas. The first is motivation—does the FCC now have the will to engage in careful cost examination of BOC operations? As the FCC has learned in other circumstances,[42] this is an awkward argument to make. The case that the future will be different requires a concession that the FCC was not inclined to do its job in the past.

The second, less embarrassing area in which changed circumstances might inspire confidence in accounting controls is that new learning makes accounting more effective. For example, one could argue that advances in personal computing have reduced data reporting and processing costs to an extent sufficient to make detectable cross-subsidization that in a pen-and-paper world would escape the eyes of FCC or state regulatory accountants. To my knowledge, such an argument has not been made. The "technical advance" on which new expectations of effective accounting controls are based seems to be a requirement that the BOCs produce "cost manuals" and adopt "rules for transactions ... designed to protect against cross-subsidization."[43] It is difficult to generate much enthusiasm for the proposition that new manuals and rules will do a job that past experience showed was impossible.

---

[41] Brock, G., *The Telecommunications Industry* (Cambridge, MA: Harvard University Press, 1980): 207-210.

[42] An argument the FCC has employed in favor of price caps is that conventional rate-of-return regulation induces the regulated firm to operate inefficiently. Evidence of such inefficiency, though, would be evidence that the FCC did not meet its statutory responsibilities.

b. The Incremental Costs Standard—ONA Reprise

Assuming cost data are obtainable, the next question is what to do with them. To protect against the possibility that cross-subsidies from ratepayers may end up covering below-cost pricing in other markets by a regulated firm, an incremental cost ("IC") standard has been proposed.[44] The IC test is that the prices charged by a regulated firm in an unregulated market, say by an LEC for information services, must generate revenues covering the added total cost to the regulated firm brought about by the decision to enter the information services market. For example, if providing electronic mail costs an LEC $15 million more than it costs just to provide local exchange service, its prices should be enough to generate at least $15 million in revenues.

The appeal of this test, especially when compared to arbitrary "fully distributed cost" pricing rules, is that a service covering its IC is not leaving the ratepayers any worse off. There is no shortfall between revenues and costs that the ratepayers are being asked to cover.[45] The IC test, though, addresses only predation. It fails to police against cost misallocation, which is the heart of the cross-subsidy problem. A cross-subsidizing firm need not be selling at a price below costs, or even below its competitors costs. The purpose of cross-subsidization is not predation but evasion of regulation. Cross-subsidization is the end and entering other markets is the means, not *vice versa*.

---

[43] Triennial Review Order at 128-31.

[44] Breyer, S., *Regulation and Its Reform* (Cambridge, MA: Harvard University Press, 1982): 52-54; Temin, *supra* n. 24 at 125.

A potentially more important flaw in the IC test has been pointed out, originally by Baseman.[46] If a regulated firm knows that its offerings in competitive markets will be judged by the IC test, it can design its production facilities to maximize the extent to which costs are "common," i.e., cannot be attributed to either local exchange or the information service,[47] and minimize the incremental costs of providing information service. A service passing an *ex post* IC test nonetheless may be inefficiently provided *ex ante*, if "incremental costs" are artificially low because of a strategically motivated technological choice. This generates a cross-subsidy, but through inefficient plant design rather than improper accounting. Under an IC test, costs shifted from being "incremental" to being "common" through design choices effectively get passed on to the ratepayers.[48]

---

[45] Breyer, *Id.*; Baumol, W. and Bradford, W., "Optimal Departures From Marginal Cost Pricing," *American Economic Review* 60 (1970): 265-83; Faulhaber, G., "Cross-Subsidization: Pricing in Public Enterprises," *American Economic Review* 65 (1975): 966-77; Breautigam, R., An Analysis of Fully Distributed Cost Pricing in Regulated Industries," *Bell Journal of Economics* 11 (1980): 182-96; Sharkey, W., *The Theory of Natural Monopoly* (Cambridge: Cambridge University Press, 1982): 107.

[46] Baseman, K., "Open Entry and Cross subsidization in Regulated Markets," in Fromm, G., *Studies in Public Regulation* (Cambridge, MA: MIT Press, 1981): 329-60. Baumol ("Comment," in Fromm, G., *Studies in Public Regulation* (Cambridge, MA: MIT Press, 1981): 361-64) criticizes Baseman's argument on the grounds that such an inefficiently designed facility could survive only if the firm were protected from entry by law. With regard to local exchange service, the size of the irreversible investment needed to enter local exchange service, and the ability of the incumbent to respond, suggest that large scale entry is not something LECs should fear, even absent local franchising.

[47] A cost "attributable" to a service is one that would not have been incurred had that service not been provided.

For information services, such costs may be shifted by designing inordinately large central office switches, in which the components perform both local exchange and information services. A more efficient design might entail separate components, but a large switch might be profitable if the lion's share of its costs were allocated to the ratepayers under an IC standard for information service offerings. If technology permits a range of component designs, policing against cross-subsidy involves not just examining the books, but in second-guessing the technical design of local exchange networks.

Such second-guessing might be viewed as beyond the competence of the regulator, but one current policy initiative connects the FCC intimately with the fine points of network design—ONA. Ironically, a policy of limited effectiveness against its intended target, discrimination, may turn out to be an effective policy against cross-subsidization. As noted above, the goal of ONA is to break the local exchange down as far as possible into separately obtainable and tariffed service elements. In Peter Huber's words,

> Under [ONA], LEC switches will perform only the most basic switching functions. The ONA concept recognizes that many services can be

---

[48] Baseman (*supra* n. 46 at 330-31) provides a numerical example; the theoretical effects of this cross-subsidy tactic are set out in Brennan, *supra* n. 10. That analysis and the following example in this paper were suggested to me by a presentation by John Panzar at the 1986 Telecommunications Policy Research Conference.

> provided from a limited set of elemental switch capabilities, called primitives.[49]

If the FCC succeeds in implementing ONA as Huber envisions, it may be impossible for a BOC to design a digital central office switch to recast "incremental" information service costs as common to exchange services. Note, though, that not all observers believe this goal is currently close to being realized.[50]

C. *Antitrust Literature Predation Tests*

Antitrust policy experts have been debating for some time whether firms would ever predate and how a judge in an antitrust case could tell whether predation took place. The criteria for predation developed in this literature might be useful for policing anticompetitive conduct by LECs in information services markets. Areeda and Turner provided the best known of these tests, in which a price is deemed predatory if it fails to cover its "average variable costs" of production.[51] For practical purposes, this is indistinguishable from the IC test described above and shares those virtues and vices.

---

[49] Huber Report, 6.23-24.

[50] Hatfield Associates, *supra* n. 40; Triennial Review Order at 138-42.

[51] Areeda, P. and Turner, D., "Predatory Pricing and Related Practices Under Section 2 of the Sherman Act," *Harvard Law Review* 88 (1975): 697-733. It probably should be noted that with any economic or legal test, assorted marginal qualifications may apply.

The follow on literature to Areeda and Turner is vast;[52] only a couple of the more potentially useful predatory tests will be described here. Baumol has proposed that a firm's prices not be found predatory if they do not rise upon the exit of a competitor.[53] The usual theory of predation is that it involves unprofitable short-run pricing to drive out competition; the profits come only later when the predator can charge the monopoly price. If post-exit price increases are grounds for antitrust liability, though, they would be deterred, removing the opportunity to cover the losses incurred in setting prices below cost. Williamson examines quantity rather than price, claiming that a firm not be permitted to increase output upon entry of competitors.[54] Absent predatory motives, the least a firm would do upon entry would be to maintain output; output reductions would normally be expected as new entry causes price to fall. Increasing output in response to entry would, on this account, signal predatory intent.

Both of these rules share the virtue that examination of costs is not needed to infer predatory intent. Unfortunately, their usefulness in the information services context is limited at best. An LEC may be able to deter entry into information service markets by the threat of cross-subsidization or discrimination, and Baumol's or Williamson's rules would apply if such conduct was carried out in response to entry by unaffiliated ISPs. But the prime motive for these tactics is not strategic predation but regulatory evasion. Cross-subsidization may be profitable without inducing the exit of competing ISPs, or even causing price to fall in the subsidized information services market. Their applicability to discrimination is even more remote, since the effect of discrimination is to raise prices in information service markets, not cause them to fall.

---

[52] For an excellent survey, see Ordover, J. and Saloner, G., "Predation, Monopolization, and Antitrust," Department of Economics, New York University, 1987.

[53] Baumol, W., "Quasi-Permanence of Price Reductions: A Policy For Prevention of Predatory Pricing," *Yale Law Journal* 89 (1979): 1-26. A conceptually similar predation test, to deter investments in productive capacity or innovation profitable only because a firm exits and would not reenter if the price subsequently rose, is proposed in Ordover, J. and Willig, R., "An Economic Definition of Predation: Pricing and Product Innovation," *Yale Law Journal* 91 (1981): 8-53; their test is criticized in Schwartz, M. "Welfare Effects of Exit-Inducing Innovations," *EPO Discussion Paper 84-12*, Antitrust Division, U.S. Department of Justice (1984).

[54] Williamson, O., "Predatory Pricing: A Strategic and Welfare Analysis," *Yale Law Journal* 87 (1977): 284-340.

WHY HAVEN'T WE SEEN THE INFORMATION SERVICES REVOLUTION?

The good news is that cross-subsidization and discrimination require a special set of regulatory and market circumstances sufficient to make them less likely than a reading of *U.S. v. AT&T* might indicate. The bad news is that *ex post* antitrust enforcement or regulatory reforms cannot be expected to eliminate whatever risk remains. To the extent this question is valid, it might be worth entertaining some hypotheses regarding the main justification for dropping the prohibition of BOC entry into information service markets. The main argument is that the U.S. is falling steadily behind in the "information age," exemplified by the widespread use of the Minitel system in France and the absence of a similar system here. We would have such a system here, by implication, because economies of scope render the BOCs uniquely able and willing to provide these services to the American consumer, but they are kept by edict from doing so.

As recognized above, economies of scope are a powerful justification for letting a regulated firm into an unregulated market despite the risks posed by cross-subsidization and discrimination. But, as also noted, economies of scope between local exchange service and information services are limited, especially if bred out of the network by ONA. Thus, it may be worth while to review other hypotheses to explain why we are not all ensconced before data terminals.

- *Lack of demand.* As with direct broadcast satellites, videodiscs, and to a large extent personal computers (except for business or education), the American public may not be all that interested in many of the lamented electronic data services. It may be unreasonable to expect much demand for computerized information services, when setting the timer on a video cassette recorder is regarded as worthy of an honorary degree. Note also that to some extent Minitel was forced upon the French market by a policy prohibiting the distribution of printed telephone directories.

- *Poor MFJ drafting.* It is certainly possible that exchange telecommunications and information services were poorly defined and interpreted at the outset of the divestiture. While permitted "exchange telecommunications" appears by definition to permit wide latitude in what BOCs may offer, they may have been prohibited from providing some information services for which they possess considerable economies of scope. Some of these services ought to have been considered exchange services in the first place; protocol conversion may be a good example. This seems the most reasonable interpretation of the court's current view toward letting the BOCs into information transmission while keeping them out of content generation.

- *Inherent inefficiency of regulated firms.* Under conventional regulation, a regulated firm earns no profits regardless of what it does. While this may mean the firm has no incentive to monopolize markets, it has no incentive to respond efficiently to changes in technology or consumer demand. Here, a regulated BOC has no incentive to facilitate the introduction of information services, even if it were efficient, since it gets no profit at the margin for doing so. An analogous circumstance was the requirement that the post-divestiture BOCs provide equal access to interexchange carriers. According to the theory of the divestiture, all incentives to discriminate in favor of AT&T should have been eliminated. That may have been true, but cost-of-service regulation still gives no incentive to the BOCs to cooperate with the other interexchange carriers, rendering explicit orders necessary. The absence of similar orders to accommodate ISPs may have deterred the onset of some information services; more regulatory activism in this area may not be a bad idea. For example, it may be sound policy to order the BOCs to adopt ONA, even if they were not permitted to enter information service markets.

- *Pessimistic expectations by ISPs.* As noted in the introduction, the policy trend is to let the BOCs into information service markets. We have also shown, though, that cross-subsidization and discrimination are still risks faced by those who would compete with BOCs. To the extent entry into an information services markets is a long-term project entailing extensive irreversible investments, a potential ISP may be reluctant to enter if the ISP foresees entry by a firm with the incentive and ability to cross-subsidize or discriminate against the ISP.

- *BOC refusal to cooperate.* The BOCs may have been anticipating ability to enter as well. If so, it is at least possible that they may be "discriminating" in advance, e.g., refusing to cooperate with the information services industry to set standards or otherwise facilitate development of that market, before they have a chance to enter themselves. Since common standards for information transmission and terminal interfaces are an important component of the success of services such as Minitel, the absence of such standards here may noteworthy.

- *Different regulatory structures.* The emphasis in the policy debate over information services has been on the degree to which LECs are permitted to operate in unregulated information services markets. What has not been questioned is the tradition, going back to "Computer II," that services other than basic local transmission and switching be unregulated. In "Computer III," the FCC considered but later rejected replacing this "definitional" approach with one more closely tied to the economic characteristics of the service. Had the change in approach been adopted, the door might have been opened to identifying services that the BOCs ought to be able to provide because of economies of scope, but that could remain regulated in some fashion to take away some of the incentive to cross-subsidize or discriminate. Profits from such services, for example, could be earmarked periodically to defray the cost of local exchange service. If such a class of services could be specified under U.S. regulation, there might not be the severe "either/or" dichotomy faced by U.S. policy makers—either the BOCs can't enter, or they can operate without any regulation. Foreign regulators may not be squeamish about broadening regulation, thus they may not view expanding the service domain of their telephone companies as radical, traumatic, or risky.

Structural policies, and substitutes such as antitrust enforcement or regulatory reform, warrant the reexamination they have been getting in recent years We might also reexamine policies that, without a lot of consideration, force risks of cross-subsidization and discrimination whenever there are economies of scope to be reaped or industrial policies to be implemented.

THE EFFECTS OF HOUSEHOLD CHARACTERISTICS ON TELEPHONE USAGE AND CLASS OF SERVICE CHOICE

Alan J. Cox

Graduate School of Business Administration
University of California at Berkeley
Berkeley, CA, 94720 Tel: (415) 547–8481

1) Introduction

A pervasive problem of regulation of public utilities that are natural monopolies has been the allocation of common fixed costs amongst different users and types of service. In the telephone industry the problem of how to price subscriber access and usage is receiving a great deal of attention. The problems of common costs have caused some debate since the implementation of the Modified Final Judgment (the MFJ) which resulted in the divestiture of the local Bell operating companies by AT&T. A large component of these fixed costs (or non–traffic–sensitive costs) has been in the provision of local access to the telephone system, both the wires emanating from the telephone companies local switching office and the equipment necessary to switch calls to other users or to other networks. These costs have been paid for through a variety of instruments as we shall see below.

The theory of public utility pricing indicates some solutions for these problems. Absent any political consideration, the preferred solution would be to price all goods at marginal cost and make up the resulting deficit from funds raised by a non–distorting tax. More realistically, in the U.S. utilities are constrained to break even. Second best solutions to this constrained optimization problem have been developed. These include uniform Ramsey pricing, (described in Baumol and Bradford, 1970) by which a commodity produced by the regulated utility is sold at different prices in different markets. In each market, the markup over cost as a percent of the price is proportional to the inverse of the elasticity of demand in each market. That is, those customers whose usage is relatively price insensitive will be charged more than those customers who will, for instance, consume a good deal less as a result of a price increase.

Another second–best rate structure, two part tariffs, allows for different customers in the same market to be charged different prices. These tariffs are more generally referred to as non–linear tariffs. A non–linear tariff can consist of a fixed "entrance fee" with a constant price for services actually used. Classic examples of this situation in the economics literature include, i) amusement parks which charge an entrance fee and for each ride, ii) the rental of a photocopying machine with an additional charge for each copy made and iii) membership on a club for which there is an initial entrance fee and additional fees for each usage of the facilities. Another type of non–linear tariff is that by one rate for the first (say) 100 kWh of electricity consumed and then a lower rate for each additional kWh consumed. Non–linear tariffs are discussed in Brown and Sibley (1986), Leland and Meyer (1976) Goldman, Leland and Sibley (1984), Oi(1971), Schmalensee (1981), and Ng and Weisser(1974). Two part tariffs with quality discrimination are discussed by Oren *et al.* (1983).

For Ramsey prices to be put into effect the utility must be able to discriminate between customers of different types. Discrimination of this sort is not usually practical or, indeed, legal. Non–uniform tariffs, on the other hand, can be designed so that customers choose their own price schedule. Properly designed rates will induce heavier users, who value the service a great deal, to pay high entrance fees for the use of the product at a low price. In the parlance of economics, these users will be described as having a high "consumer's surplus"; they value the use of the commodity well above the cost of usage. Rates can also be offered for smaller users who use the product very little and would be willing to pay higher unit costs but only for a low entry fee.

But determining who will choose what rates and what the revenue impact will be of such offerings requires knowledge about the distribution of tastes for calling and subscriber access. In the absence of that information telephone companies may, for instance, set access fees so high that, even with very low usage fees, few will take the service. More subtle pricing strategies will require an ability to forecast how many persons will switch into a class of service if the monthly access fee is raised in return for lower usage fees. With information on, properly designed non–uniform rates may reduce the need to cross–subsidize the product while still increasing welfare over that possible with a set of uniform Ramsey prices.(Willig, 1978, Brown and Sibley, 1986, pp.83 ff.)

In the case of telephone pricing, unfortunately, little work has been done in estimating demand at the household level. Such work is necessary in order to determine the distribution of tastes and of demand levels amongst the population of potential telephone users. Only then will it be possible to design rate structures that will come closer to meeting goals of high telephone availability with minimal, if any, cross–subsidy. Such work is also necessary to determine whether there are any particular sectors of society that will be particularly hurt by such rate changes and whether there are any public policies that could be undertaken to mitigate that impact.

This paper starts with a very brief review of the historical background of the problem of fixed costs in the telephone industry and then a brief overview of the theory of two part tariffs. That is followed by a brief description of the rate structure facing the potential telephone customer. The rate structure for telephone services introduces complications in the conventional economic analysis of demand. Technically the problems are described as a budget set with non–convexities and non–linearities. These problems are discussed in that section.

Section 5 follows with a description of the complex utility optimizing problem that result from the rate structure. This section also contains a discussion of the evidence on class of service choice. More explicitly we discuss in that section whether there is any evidence that telephone users are making class of service choices on the basis of anything that is measurable, such as cost–minimizing or risk–averse behavior. This is important since many argue that, for a variety of reasons, class of service choice is either totally random or is made on the basis of preferences that cannot be measured. Next, section 6 discusses the desirable features of a utility function for telephone usage. Such a utility function is laid out in Section 7.

This paper lays some of the groundwork for the analysis of a new data base of over ten thousand telephone users, including information on their class of service option, detailed usage data, and demographic information on the number of household members, ages of all employed persons, income, occupations, education and ethnic background. References will frequently be made to that data base.

2) Historical Background

Before the implementation of the MFJ parts of the non–traffic sensitive costs of local service were covered through "separations" process. Each of the AT&T subsidiary operating companies received a portion of the revenue generated from long distance service. The size of the payment to any particular local operating company depended upon the amount of long distance traffic that had been generated from that utility. The funds transferred in this way were used to defray the costs of maintaining the access of telephone customers to local switching offices.

The divestiture by AT&T of its local operating companies eliminated a large part of the funding of these direct cross–subsidies. Some subsidy could still be derived from more distant calls placed within the local monopoly franchise. Higher rates for business customers could also be used as a source of additional funding to subsidize local residential access. In addition to these sources of cross–subsidy, the Federal Communication Commission (the FCC) attempted to have a significant proportion of the fixed costs borne by a fee levied on each telephone customer. The initial proposal was for a fee of $2.00 per month. Even that low proposed fee, however, was large enough to attract the attention of Congress. Legislation passed there rolled back the fee to $1.00 a month.

In response to this political pressure the FCC (for interstate calls) and the state regulatory commissions (for intra–state calls) required the interexchange carriers (IXCs) to pay the Bell Operating companies (BOCs) an "Access Fee" or common carrier line charge. This, in fact, was merely an extension of the old separations process. The access fee that an IXC was required to pay varied with volume of calls that the BOC's passed from its customers to the IXC. Nearly half of the price of a long–distance telephone call placed through AT&T within California in 1985 was due to the need to raise funds to cover these access fees.

Very large telephone customers had a strong incentive to avoid BOCs altogether and pass their long distance calls directly on to the inter–exchange carrier. This could be done by merely transmitting calls over a microwave system from the roof of the telephone customer to the "point of presence" of the interexchange carrier. Since the amount of the access fee paid to a BOC depended upon the number of calls that the BOC transmitted to the IXC, bypassing reduced the cost of access and allowed the IXC to pass those savings on to the by–passing customers. While only a small number of customers found it economic to undertake bypass they represented a significant amount of revenue for their respective BOCs. Bell Atlantic, for instance, estimated that by–pass would result in a loss of 20 per cent of the access charges it would otherwise earn by passing calls on to IXCs. (Jackson and Rohlfs, 1984).

Technological improvements and other distortions in the price structure for business telephone service induced further bypass of the local public network. An example of the former includes switching capacity owned by the business customer. Services provided by the telephone company (Telco) such as intraoffice calling, call forwarding and instant redialing could be provided by customer–premise switching equipment (i.e. customer owned equipment). Pricing distortions resulted in dedicated lines being provided by the local telephone companies at a lower price than switched access. Large customers could use these dedicated lines to access their branch offices or to the Point of Presence of their inter–exchange carrier. The lost revenue to all local operating companies in 1984 from all sources of bypass was estimated to be $230 million, with the 1987 figure forecasted to be $770 million.(Fischetti, 1985, p.80)

Other developments have also made unsustainable some of the cross–subsidy of services in the telephone industry. For instance, some PUCs have expressed an increased willingness to permit resale of such services as WATS, thus allowing smaller users to band together and avail themselves of the benefits of quantity discounts.

The phenomenon of by–pass eventually forced the imposition of greater monthly fees paid by individual subscribers allowing some decrease in the access fees billed directly to the Inter–exchange carrier(IXC). In 1986 the fees was raised to $2.00 per month. By mid–1987 they had reached $3.50 per month.

The result of this and other developments is to increase the amount of revenue that must be recovered from the residential sector and from intraLATA toll calling. Before the AT&T divestiture, about 25 per cent of calls were made within the boundaries of what came to be the Local Access and Transport Areas (LATAs) and are, therefore, part of a BOCs monopoly franchise. The longer of the calls (called Measured Toll Service, or MTS) placed within a LATA franchise have, in general, been priced well above marginal cost, with the excess revenue being used to hold down local rates. However, it seems likely that the same pressures that led to divestiture will also lead to competition being introduced to the intraLATA market (Kahn testimony, California PUC, application numbers, 82–11–07, 83–01–21, 83–04–02, also Huber, 1987). The interLATA carriers all have points of presence (POPs) located throughout LATAs all over the country. Some intraLATA calls are already being handled by the IXCs, though the CPUC takes the view that this is just incidental rather than a deliberate effort by the IXCs to enhance their revenues. Again, a price structure that would allow for some cross–subsidy of local access from intraLATA toll calling would become unsustainable. Additional pressure on cross subsidies will arise in the future as new technologies allow the introduction of new services such as voice–mail and software–defined networks. Companies other than the local monopoly BOCs will continue to attempt to get rulings and legislation that will allow them to compete with the BOCs to provide these enhanced services.

It should be noted that it is not clear that a subsidy to access is actually required. The case has been made (Cornell and Noll) that the subsidy to local access that is imputed from accounting data does not exist in economic terms. Providing access through the utilization of the new technologies may be less costly than the traditional wire loop dedicated to each line. Even if there were no changes in technology, there is some evidence that the construction of the local loop by the regulated telephone monopoly was inefficient. Excess capacity of 35 per cent is not uncommon in many regions of the country though that fact alone is not an indicator of inefficiency. There is also the possibility that locations could be more efficiently wired for many services at the same time. CATV systems, for instance, could be installed at the same time as telephone loops and could substitute entirely for some of the services provided by telephone access.

All these considerations have led some to conclude that there is room for increased competion and deregulation in the provision of even local telephone services. However, there is little indication that the *status quo* is likely to change in the near future. Judge Harold Green, the presiding judge on the AT&T anti–trust case and the author of the Modified Final Judgment, continues to supervise the evolution of the structure of the telephone industry in the United States. His most recent rulings indicate that he has no intention of permitting any major inroads to the local monopoly of the BOCs in the near future.(Telecommunications Reports, Sept. 14, 1987). At the same time he refused to allow the BOCs to provide inter–LATA telephone service or to provide customer premise equipment. He based his opinion on his finding that "objective economic conditions entirely preclude the provision of local distribution functions at a lower or equal economic cost than could the established local exchange carrier. There is likewise no indication that the (BOCs') natural monopolies have been eroded by technological change."

Maintenance of the *status quo* includes continued subsidy of local access by charging for toll calls and business services at a rate greater than cost. In California this is both implicitly and explicitly done. One explicit cross–subsidy is a Public Utility Commission (PUC) mandated tax on MTS calling that is earmarked to underwrite the cost of providing lifeline service. Thus the current structure of telephone rates and cross–subsidies can be seen to be generally acceptable to those with the political responsibility for supervising rate structures. In order to maintain the relative calm of the telephone regulatory process some PUCs are entering into "social contracts" with the Bell Operating Companies. Under these social contracts the BOCs undertake to maintain the current residential rate structure for a set number of years. During that period they are allowed to price business services in any manner that they see fit to compete against the various by–pass technologies. They are also allowed to keep any savings that might result from improved management or the incorporation of new technologies into the system.

Forms of "social contract" regulation have been actively proposed or enacted in Alabama, California, Nebraska and Vermont. (See Telecommunications Reports, 1985 and 1986.) The Nebraska legislature enacted a law (LB 835) that has since been challenged before the supreme court by, amongst others, the Public Service Commission (PSC). Under the law a Local Exchange Carrier could raise rates on basic local service without a review, though the PSC could initiate a review of any rate increase of more than 10 percent and consumers could petition for a review of any increase. PSC–initiated reviews are only to ensure that services are not being priced below costs; services priced above cost will not be trimmed back if they were above costs at the time the legislation was enacted.

As another example of social contract legislation, Pacific Bell has proposed that residential prices be frozen in nominal terms until 1990, with slight adjustments through 1995. At the same time Pacific Bell proposed that it be allowed the introduction and/or pricing of other product service offerings particularly in the business sector. (See testimony of Lee Camp to the CPUC, application 85–01–034.) The strategy for Pacific Bell that these tactics indicate is to generate enough revenue from other products in which it has a comparative advantage in order to remain competitive in the residential sector, while maintaining a policy of maintaining near–universal access to the telephone system. It is able to maintain current low access prices and low costs for local service through network access fees and MTS service. These represented 32 and 21 percent, respectively, of Pacific Telesis' revenues in 1985.

It may be that these revenue goals are more appropriately achieved through other pricing arrangements. Non–linear tariffs may exist that would meet the objective of maintaining a high degree of access and minimizing the degree to which local access needs to be subsidized. It is the objective of this research effort to determine the parameters of a utility function and the effects of demographic variables on those parameters. Once demand functions have been estimated for telephone service the information can be used to examine the effect of a variety of rate structures in the residential sector. Specific issues that can be addressed include the possibility of increasing revenue by altering the time–of–day structure of rates and the question of which segments of the residential sector will be made worse off or better off by changes in the rate structure.

3) The Theory of Non–linear Tariffs[1]

As mentioned in the introduction, a regulated firm with large fixed costs can recover its fixed costs if it can discriminate amongst its identifiable markets. Welfare maximization, subject to the break–even constraint, can be achieved with rates set so that customers in each market pay a price proportional to the price elasticity of demand in that market. Explicitly, the result is that the mark–up of price over marginal cost is $\frac{P_i - C_i}{P_i} = \frac{\lambda}{\epsilon_i}$ where P_i is the price charged in market i, C_i is the marginal cost of supplying market i, λ is the Lagrangian multiplier on the break–even constraint and ϵ_i is the own–price elasticity of demand in market i. (Baumol and Bradford (1970) provide the complete derivation of this result.) These results have been extended by Rohlfs (1979) to take into account the cross–elasticities of demand for the different goods a multiproduct monopoly may have.

---

[1] Brown and Sibley (1986) and Sandler and Tschirhart (1980) provide reviews of two part tariffs.

Welfare can be increased over that possible from uniform Ramsey pricing with non–uniform prices in many instances. Coase (1946) suggested that each buyer of the goods provided by the natural monopoly be charged an entrance fee equal to the average cost of providing that person with access and that the per unit price of consumption be equal to the marginal cost of providing the good. Thus there would be no distortions in the person's decisions to consume since price would equal marginal cost. Potential buyers would decide whether to join the club by merely comparing the entrance fee with their consumer surpluses at the usage price. Willig (1978) shows that for any uniform price scheme where price exceeds marginal cost there is a non–uniform price schedule that Pareto–dominates it.

The problem with the Coase formulation is that if the access fee is set too high consumers might be priced out of the market even though they might make a significant contribution to covering fixed costs. While a monopoly could try and identify such individuals and offer them lower entrance fees, such activity is expensive and possibly illegal. It is also unnecessary, since two part tariffs can be designed so that consumers segregate themselves into the tariff structure that suits them best.

Conditions that optimal two–part tariffs must satisfy have been derived by Oi

(1971), Ng and Weisser (1974), Schmalensee (1981), Phillips and Roberts (1985), Brander and Spencer (1983) and Sherman and Visscher(1982). These results depend on the behavior of the "marginal" consumer in a class or market to that of the average consumer in that class. The marginal consumer is the person whose consumer surplus is just equal to the entrance fee. His or her consumption can be greater or less than the consumption of the person with just average consumer surplus. If it is greater, Oi's results indicate that price is less than marginal cost. If it is less, then the opposite is true.

These results have been confirmed and elaborated upon by the other authors cited to take into account other cases. For example, Brander and Spencer (1983) also allow for heterogeneous subscribers to different classes of service. They show that if there are no marginal subscribers in a particular class, then that class will bear a greater portion of the fixed costs of providing service. Empirical applications of two part tariffs include Naughton (1986) who applies the results of Schmalensee (1982) and Sherman and Visscher (1982) to the U.S. electric utility industry. Mitchell (1978) estimates optimal two part tariffs for a telephone system, while Park and Mitchell (1986) evaluate the relative merits of measured service and flat rate service for a telephone utility.

An additional consideration in establishing two–part tariffs for telephone services is that of externalities. This concern dates back to the seminal paper on telephone demand by Artle and Averous (1973). The argument is straight–forward: under some fairly weak conditions, an individual's marginal rate of substitution between access to the telephone system and all other goods depends on the number of other telephone subscribers that she has access to. As a result, persons who would be sub–marginal users should be encouraged to join the network through a subsidy provided by other users. The extent of this externality is an important consideration in the current public policy debates on

access pricing; if externalities are important, then some cross–subsidy is justified. If they are very important, some argue that access fees set equal to the marginal cost of providing service could result in the collapse of the telephone system, since high prices could result in so many people leaving the network that it would significantly reduce the benefits of having a telephone to those who would otherwise stay on the network at the higher prices.

4) The Telephone Rate Structure

A general outline of the cost of access and usage of telephone service are presented in this section. The first part will be an explicit outline of the types of service that are available to the residential customer. This is followed by a discussion of the impact of the rate structure on the household budget set and the implications for estimation of the determinants of usage and choice of service. The rates described here were those that were in place during the time that data for this study was collected in 1985.

Telephone customers face a variety of rates for telephone service. Three basic classes of service (COSs) are available; flat rate service, measured service and lifeline service. These three types of service differ only in the way local calls are charged. The exact charges are found below on Table 1, "Rates".

Under flat rate service, which has a relatively high access fee, all local calls are free. Under measured service there is a charge for all local calls. However, the subscriber to measured service receives a three dollar credit against all calls. Lifeline service is available to those who certify that they meet certain means tests. They receive 60 free calls a month after which they are charged increasingly higher prices for telephone use.

At the time that the data was collected for this study some customers could make only two types of calls; local or toll (long distance)[2]. If a call was not within the local calling area it would be charged at the long distance rate. However, in two large metropolitan areas, all telephone subscribers could make calls into nearby exchanges that surrounded their local calling area at a price lower than that indicated for long–distance calls of the same distance. For each household there were two such extended calling areas, resulting in three concentric areas for local (or zone 1) calling, zone 2 calling or zone 3 calling. These are depicted in Figure 1. Calling into zone 2 and zone 3 is called ZUM calling.

The credit that subscribers on measured service receive is applied to both local calls and to ZUM calling. This has important consequences, as we will see below, since it is possible that a person who makes few local calls but $3.00 worth of ZUM calls will in fact save over $6.00/month on measured service relative to being on flat rate service.

Complicating the issue further is the variety of access fees that are charged by the telephone company. Many exchanges, particularly those located in less densely populated areas, have local calling privileges beyond what would normally constitute a local calling area. In fact, the local service area for such an exchange would include a contiguous local service area, usually a nearby larger town. Telephone subscribers in a small rural community would thus have the capability to talk with telephone subscribers in a contiguous community and be charged for that call as though it were a local call.

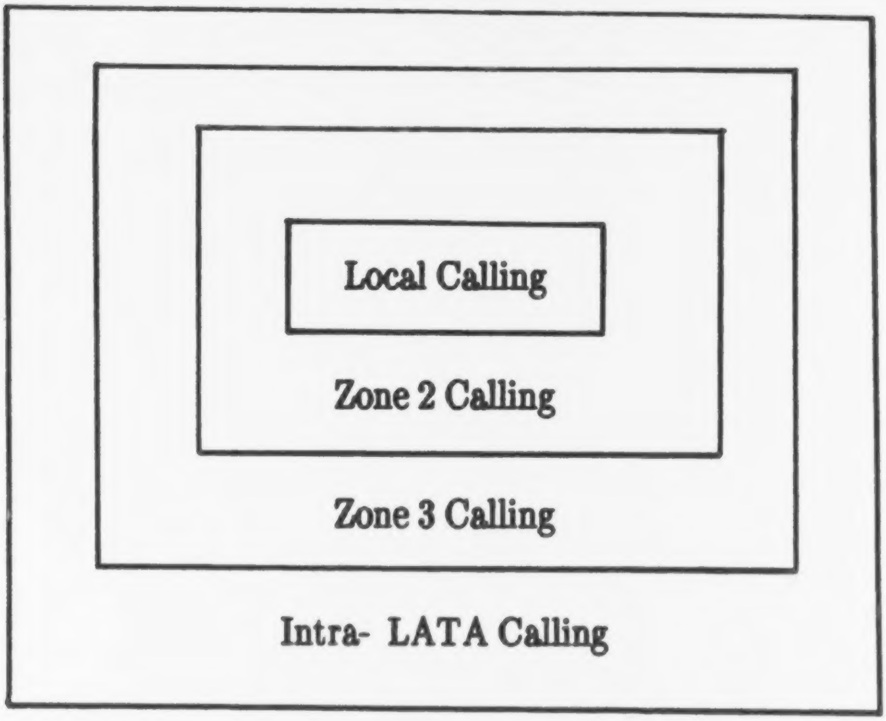
Optional Calling Plans

In addition to foreign exchange service, other "optional calling plans" were offered

---

[2] "Long distance" in this study refers to intra–LATA toll. BOCs do not have access to data on interLATA calls even though, in the case of AT&T, they still collect billing data for the IXC.

The Telephone Rate Structure


Local Calling
Zone 2 Calling
Zone 3 Calling
Intra- LATA Calling

Three TOD Periods

Fig. 1.

by the telephone company. Under these plans subscribers could reduce the cost of calling outside their local and ZUM calling areas. For this privilege, optional calling plan subscribers pay an additional monthly charge. As an example, under one such plan, "Call Bonus Circle", all calls outside the Zone 2 and 3 calling areas are discounted by 30 percent for a monthly fee of $4.75. 826 customers in this survey participated in Optional Calling Plans.

Table 1
Telephone Rate Structure Facing Survey Participants

Access Rates

Flat Rate Access	$8.25
Measured Service Access	$4.45
Lifeline Flat Rate Service	$4.45
Lifeline Measured Service	$2.30

Usage Rates

Local Calls

Flat Rate: Regular and Lifeline
All local calls free.

Measured Rate: Regular

Daytime		
	1st Min.:	4 cents
	Each Add. Min.:	1 cent
Evening		
	30 percent discount	
Night		
	60 percent discount	

A $3.00/month credit is deducted from all charges.

Lifeline

1st 60 calls:	Free
61–70 calls:	10 cents/call
Calls over 70:	15 cents/call
No TOD Discounts	

Non–Local Calls

Daytime		
Calls to Zone 2	1st Min.	8 cents
	Each Add. Min.	3 cents
Calls to Zone 3	1st Minute	10 cents
	Each Add. Min.	5 cents
Evening	30 percent discount	
Night	60 percent discount	

$3.00/month measured service credit is applied to ZUM and local charges.

5) Budget Constraints for Telephone Usage

As previously mentioned, the pricing of telephone services violates basic assumptions of conventional economic analysis. This can be most clearly seen by considering the budget allocation problem described in basic economics textbooks. Figure 2 presents the problem graphically. The horizontal axis measures the amount of good X while the vertical axis indicates the amount spent on the purchase of all other goods. The straight line shows the combinations of X and all other goods that a person of a given income can afford. The intercept of this line with the vertical axis is the amount of money that the person would spend on all other goods were he or she not to purchase any X. The intercept on the horizontal axis is the amount of X purchased if all income were spent on that commodity.

The curves shown in Figure 2 are indifference curves for a single individual. Each indifference curve maps out the combinations of X and all other goods that would give the individual a given level of satisfaction or utility. The indifference curves that are located more to the north–east indicate higher levels of utility or well–being. This is clear since more of all goods can be purchased the more in the north–east of Figure 2 the individual can place himself. Under the constraint of the budget set, maximum utility is achieved at the point at which an indifference curve is just tangent to the budget set.

One other point about conventional indifference curves should be mentioned. Note that the indifference curves never intercept the axes. This is because, under the usual assumptions of economics, individuals are never satiated with any good. If it were offered for free it would be consumed in infinite amounts.

The rate structure for telephone services results in several discontinuities and non–convexities in the budget set of consumers. Their implications for estimation are

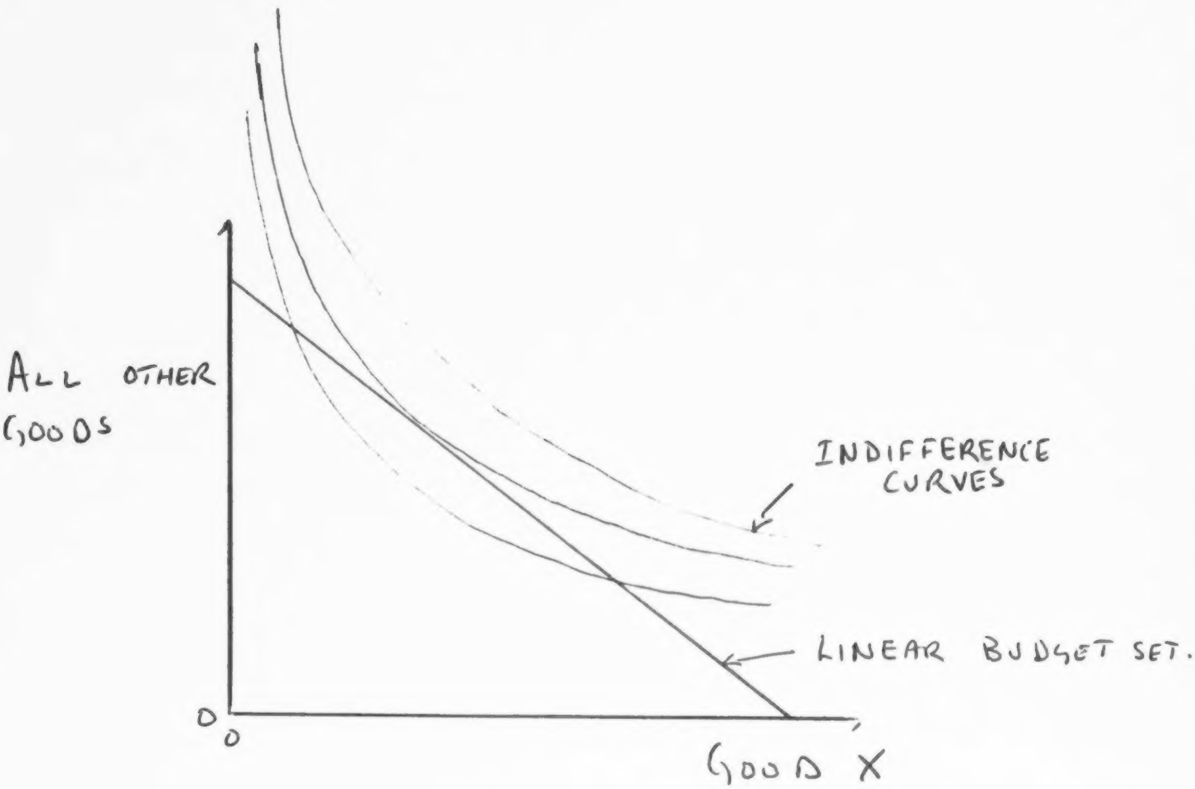

ALL OTHER GOODS
INDIFFERENCE CURVES
LINEAR BUDGET SET.
0
0
GOOD X

FIG. 2.

reviewed here. Since the shape of the budget set varies according to class of service, the effects will be discussed separately for each class of service. Appendix A provides explicit mathematical descriptions of the budget sets.

Flat Rate

A flat rate customer pays a fixed monthly amount for the privilege of making an unlimited number of local calls. Figure 3 shows the budget line for a telephone subscribing household under flat rate service. The horizontal axis represents the amount of local telephone usage (measured in calls) that can be purchased within this subscriber's household budget. The vertical axis represents the value of the Hicksian composite good, X. Y represents the households total monthly income. If p^f_{acc} is the cost of access under flat rate service then a horizontal line at $Y - p^f_{acc}$ represents the budget line for this household. Representing the budget line as flat ignores other costs of making a telephone call including the cost of not being able to receive another call and the opportunity cost of time.

Measured Rate

Under measured rate service the subscriber pays a lower charge for telephone access, pays a usage charge for each call and is allowed a credit of $3.00 against all local and ZUM calls made. Calls are priced in such a way that, for any given time of day and for any region of calling (e.g. local calling) a customer is in fact choosing the amount he will consume of two products, the number of calls he makes and the amount of time that he spends on the phone. The situation is depicted graphically in Figure 4 for two households on measured service who do not have ZUM service available to them. One household makes calls that are less than a minute in duration while the other makes calls that are on average, three minutes in duration. The first uses up its credit after 75 calls. At that point the budget line is broken and slopes downward at a rate of $.04/p_x$. The customer that makes longer calls on average has a break point at 50 calls. The budget line slopes downward at a steeper rate than that for the family that makes briefer calls. Figure 4 depicts the budgetary trade–off between calls and time spent on the telephone for constant levels of composite–good consumption that result from the rate structure. If the composite–good consumption level depicted in Figure 4 is $Y - P^{m}_{acc}$ then the line depicts the level of the credit.

The literature has tended to treat all calls within time–of–day periods and within calling zones as homogeneous goods. However, from an economic point of view, it is incorrect to treat 20 calls placed to 20 different telephones in one time–ZUM zone the same way as 20 calls placed to the same telephone in that time–ZUM zone. Increases in the price of calls in the time–ZUM zone may induce considerable economizing in the number of calls placed to the one telephone but no change in the number of calls placed to the 20 different telephones.. Train *et al.*(1987), for instance, treat calls as part of a portfolio of calls chosen from all possible combinations and zone called into. This is partly due to the aggregated nature of the data used since in both the Train *et al.* study and this study only the number of calls made to each zone and during each TOD period are known.

Variables representing the number of calls placed and duration of calling should be indexed by the telephone number being called as well as the time the call was placed. The importance of this less aggregate data can be seen by considering repeat telephone calls to the same number. A significant proportion of all calls are to the same telephone numbers. For instance, in a random sample of 10,000 ZUM 2 and ZUM 3 calls made by the customers surveyed in the study being conducted in this paper, 45% were to a number that had been called at some other time that month. For these calls there is

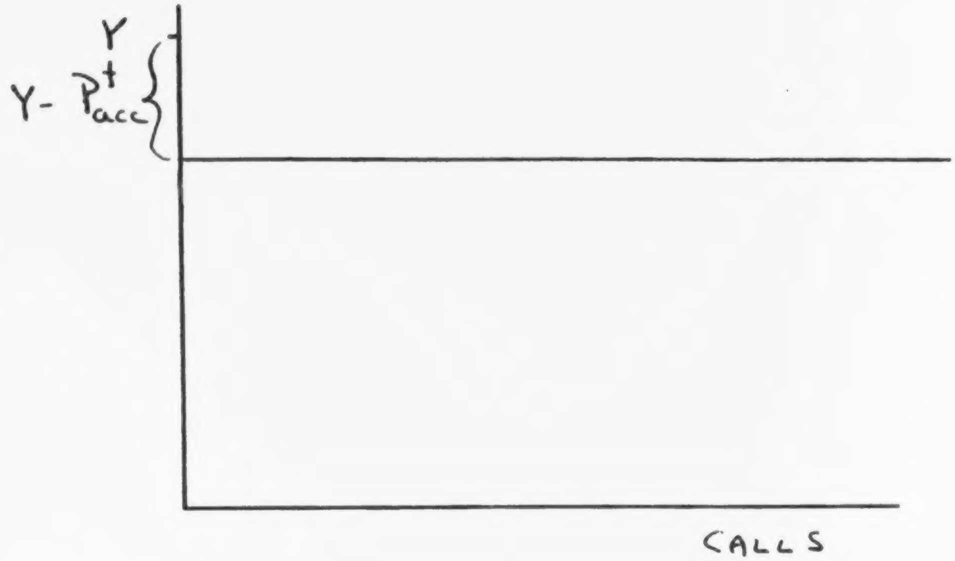


FIG. 3

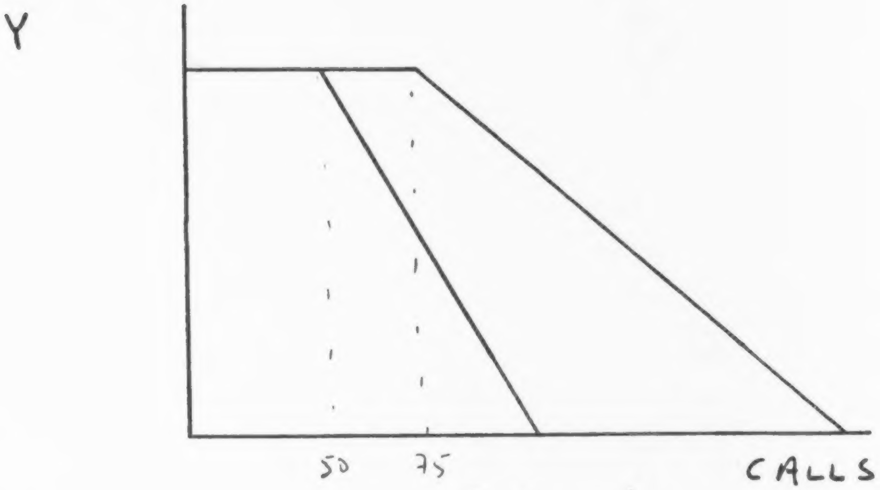


FIG. 4

some opportunity to substitute between number of calls and duration of calling. Higher fees for the first minute coupled with lower fees for each additional minute will result in fewer but longer calls being placed to these numbers. The effect of price changes will not be the same for all types of calls.

The budget sets have been depicted under simplifying assumptions. Most importantly we have assumed that calls can be made in real, rather than integer, quantities. While such an assumption may give good results for local calling under present rate structures, it may not be so for long distance calls or for calls placed in a different pricing regime. Figure 5 depicts the budget set when the discrete nature of calling is taken into account but if duration is priced as a continuously available good. Calls are charged only on the basis of duration of calls. The axis coming out of the page is measured in average duration of a call, D. For any D the length of the horizontal segment, (representing the credit available,) will be $^{C}/_{(P_D * D)}$, where P_D is the price of a call and C is the value of the credit. As one makes calls over those that would be covered by the credit, assuming that all calls are constant in duration, one gives up a discrete amount of the composite good. The height of each step, (i.e. the amount of the composite good,) is equal to P_D*D. Thus the height of each step is increasing as average duration increases and each step eventually runs into the plane of average duration and calling, where consumption of the composite commodity is zero..

Figure 6 merely serves to depict the situation where duration is priced in increments of one minute. This is the situation for virtually every telephone subscriber. The reminder of the diagrams are drawn under the assumption of continuous variables.

The situation is further complicated when ZUM calling is allowed. The $3.00 measured service credit is also applied against a subscriber's ZUM calling. In modeling

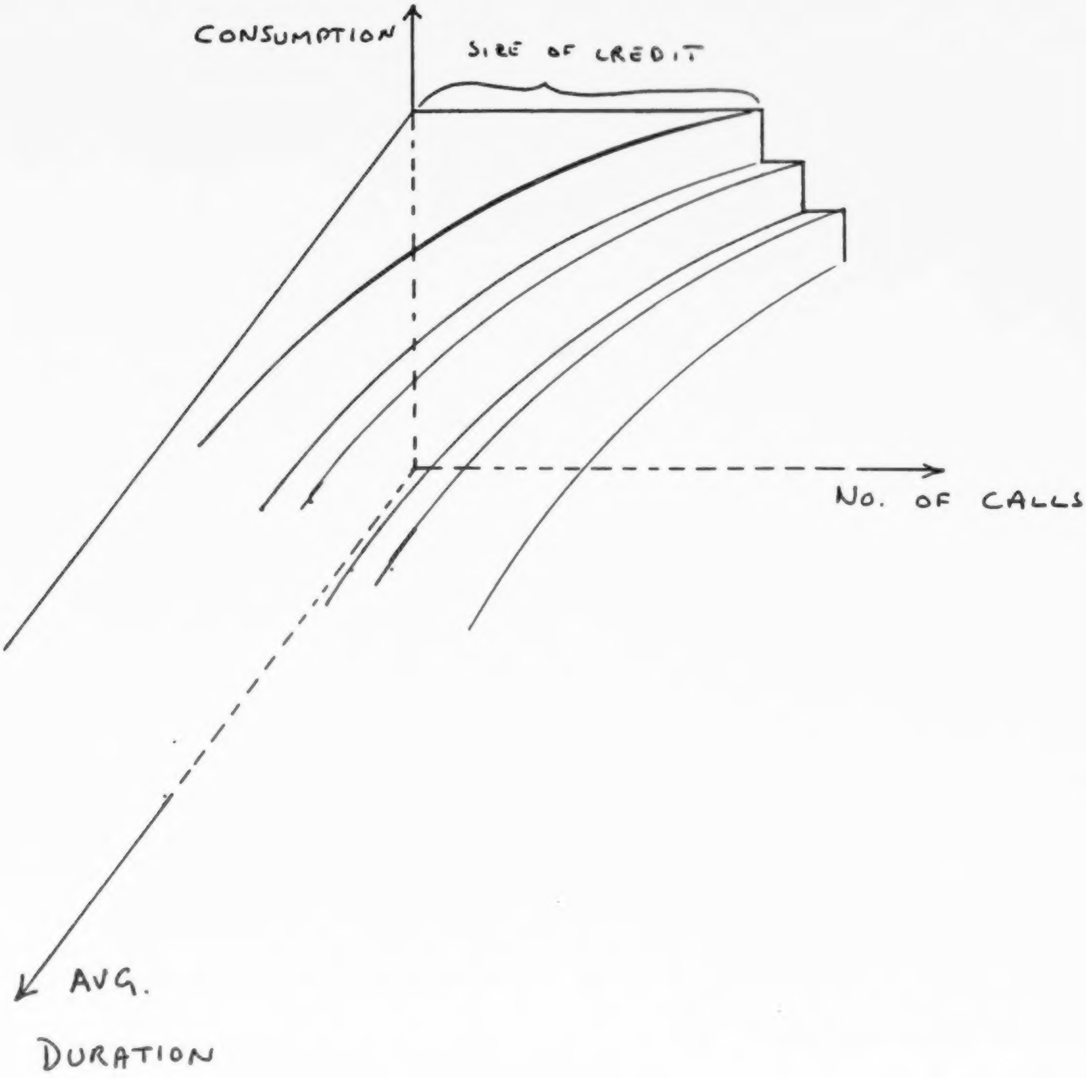

CONSUMPTION
SIZE OF CREDIT
NO. OF CALLS
AVG. DURATION

FIG. 5

21 (a)

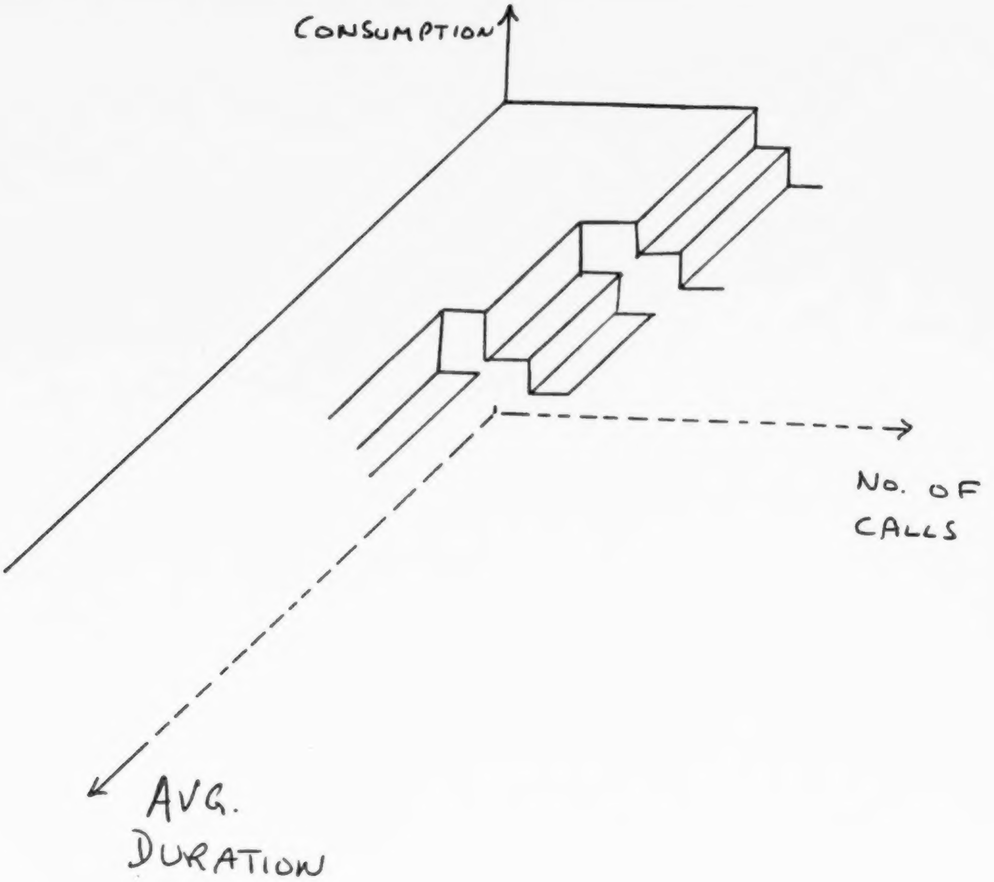

CONSUMPTION
No. OF CALLS
AVG. DURATION

FIG. 6

21 (b)

consumer's behavior it is important to know how an optimizing subscriber to telephone service allocates the credit. In some cases the credit may be large enough to induce some calling where there might otherwise have been none or a smaller amount of calling. This may have an important impact on calling behavior.

The Choice of Class of Service.

Most utility functions utilized in econometric studies of demand satisfy non–satiation assumptions. These assumptions result in an infinite demand for goods of zero price and are, therefore, clearly inappropriate for describing the behavior of telephone subscribers under flat rate service. A full income approach, as developed by Becker (1965), would allow the incorporation of the value of time and thus add an increasing penalty as the number of calls increased. Examples of the utilization of the full income approach include Gronau (1977), Graham and Green (1984), Kooreman and Kapteyn (1987) and Wales and Woodland (1977).

Figure 7 shows the budget set that is faced by a potential telephone subscriber who does not have lifeline service as an option. This caller makes three minute calls on average. The indifference curve shown is generated from an addilog direct utility function. Only two goods are included in this utility function, the number of calls and the amount of the composite good. As a result, it does not explicitly take into account the value of time of the subscriber. The problems that such a simplistic optimization results in are obvious. The graph clearly indicates that this individual would be better off going to flat rate service. Under measured service he would make about 90 calls a month (3 a day) spending a total of 4.5 hours per month on the telephone. However, according to the simplistic analysis, this caller would be better off by spending a little more money to subscribe to flat rate service and make at least 161 calls per month (over

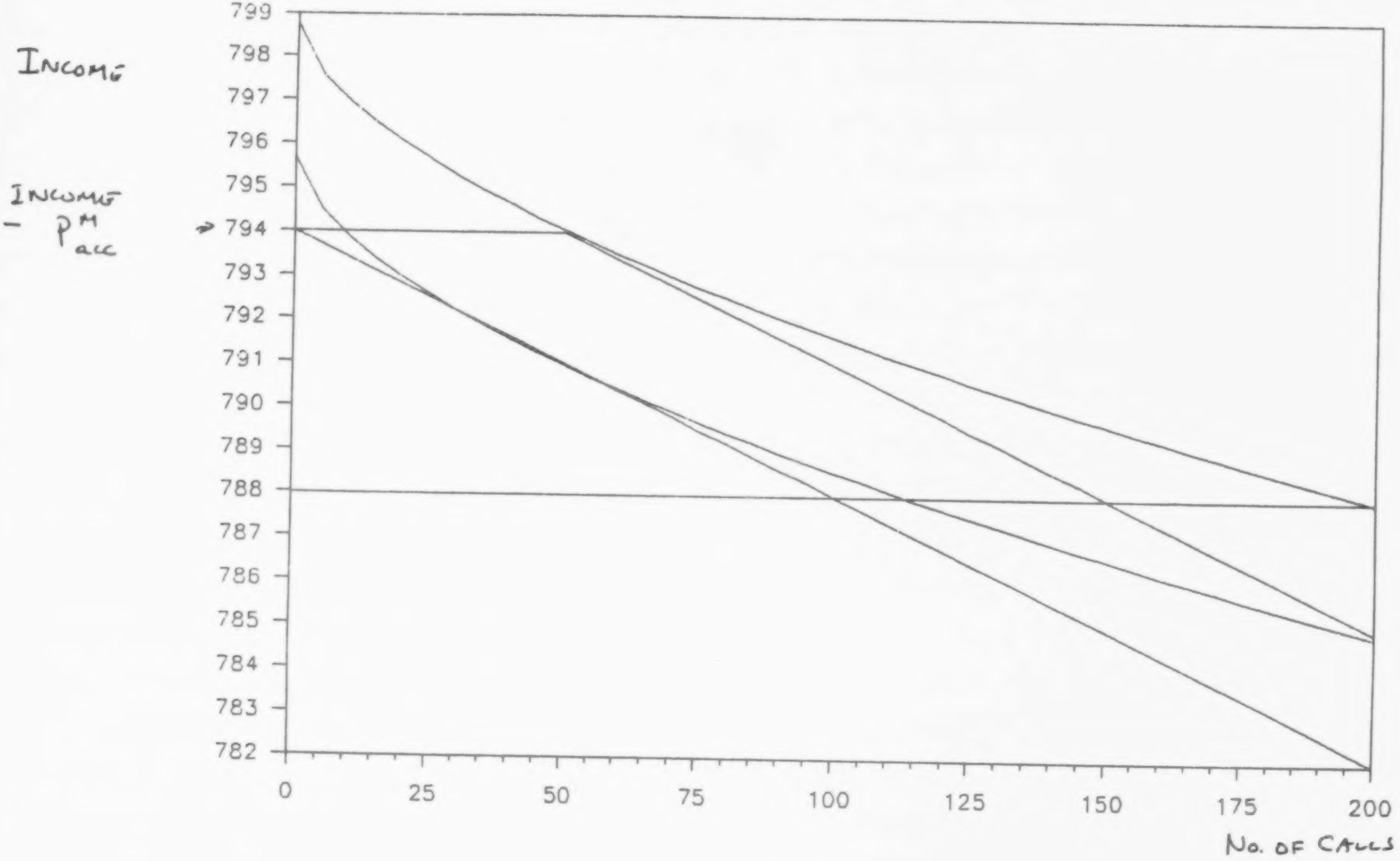

Income
Income − P^M_{acc}
→ 794
799
798
797
796
795
793
792
791
790
789
788
787
786
785
784
783
782
0
25
50
75
100
125
150
175
200
No. of Calls

FIG. 7

five calls a day). His time expenditure to make these calls would be over eight hours per month. In fact, he should make even more calls, thus achieving higher and higher levels of utility, until some binding constraint of time availability is reached. The reality of the situation is that most households make very few calls, with over 5 per cent of the households in this survey making no local calls during the first two months that their usage was monitored.

Before proceeding to the derivation of demand functions for telephone service it is worth noting briefly some of the critical issues with regard to maximizing utility when faced with such a budget constraint. Figure 8 illustrates one situation which violates conventional assumptions of demand analysis. The level sets of a utility function for a single individual can be tangent in more than one place with the budget constraint in the presence of the telephone pricing schemes outlined above. This is indicated by the indifference curves U_1 and U_2, which are consistent with the usual assumptions about utility. One of the tangency points is clearly preferable to the other. In estimating demand functions for telephone service it is necessary to both compute the local optimal and then calculate the *maximum maxorum* of those.

Figure 9 demonstrates two other difficulties resulting from the non–convex budget set. Here, again, the household has the choice between measured service and making an unlimited number of free local calls. The choice will depend upon the values of the parameters of the utility function. U_1 and U_2 here indicate possible shapes of the indifference curves for two different telephone subscribers. There will be an infinite number of critical combinations of these parameter values that will result in the utility function lying in the position depicted by U_1. More importantly, U_1 illustrates a well–known property of two part tariffs: nobody who correctly anticipates the amount of calling will intend to make the number of calls around the kink–point $\tilde{N}$. To the extent

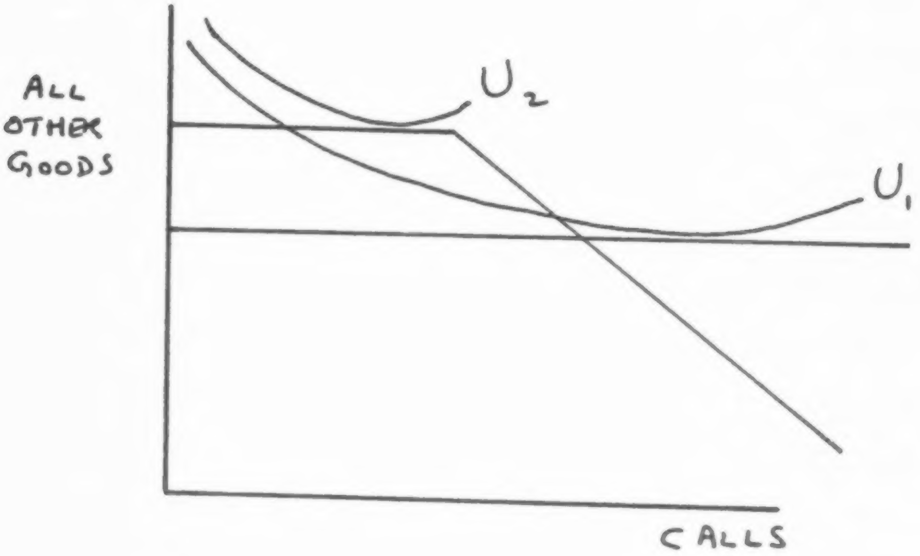

ALL OTHER GOODS
U2
U1
CALLS

Fig. 8

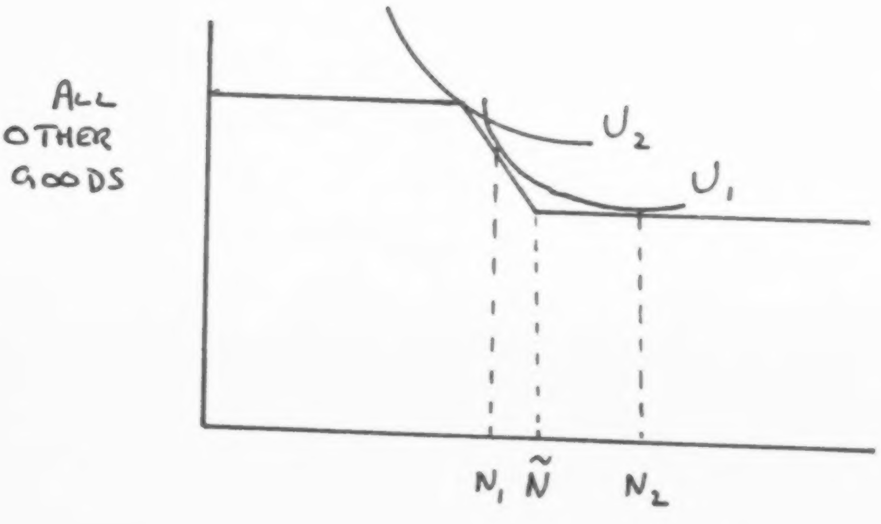

ALL OTHER GOODS
U2
U1
N1
Ñ
N2

Fig. 9

that persons are observed in that region it is necessary to model a relationship between intended number of calls and actual calls made.

There are other regions of Figure 9 where intentions are clearly at odds with actual behavior depending on class of service chosen. For instance, any household on measured service would not expect to be making more than N_1 calls and no household on flat rate service would expect to be making more than N_2 calls. (N_1 and N_2 are determined by the tangency of U_2 to the budget set.)

A second area of interest in Figure 9 is that indicated by U_2. Here the household is at a kink. Again, there are many values of the parameters of the utility function that are consistent with being at the kink. Even if the utility function had only one parameter that was allowed to vary amongst the population, there would still be a range of values of that parameter that is consistent with being at the kink.

Empirical Results on Class of Service Choice

An important question before undertaking a study of this sort is whether persons make class of service choices based on observable variables. In fact, we do find a great many telephone subscribers whose telephone usage is not consistent with their class of service choice when we consider only cost–minimizing behavior. This is particularly true for those on flat rate service. There are several reasons that strict cost–minimizing behavior may not be observed. The most obvious is that subscribers do not know or cannot understand the telephone rate structure. However, even if there were no such information problem, it is likely that many persons would still prefer to pay more in order to take flat rate service. Some survey results indicate that complete information only reduced by half the number of persons who should have been on (LMS), if they were cost–minimizers. That is, even if subscribers new that based on their telephoning behavior they could reduce their bills by subscribing to local measured service they reported themselves as preferring flat rate service. Such subscribers are probably risk averse and find the difference in the cost of access (plus whatever credits that are not available to them) worth the insurance against unexpectedly high local bills. If the latter were true we would expect to find that those subscribers on flat rate service who should be (in a cost minimizing sense) on LMS would have greater variance in their calling patterns than those on LMS who have correctly made the cost minimizing decision. Such a greater variance would indicate a greater probability that they might get an unacceptably high bill and might make it more difficult to estimate the average use of the telephone.

Others have argued that, since telephone costs are so low relative to most people's incomes, few are going to pay any attention to their telephone bills. Class of service choice will merely be a random flip of a coin. Another reason to expect such a method of choosing class of service is that subscribers are usually expected to make snap decisions on this matter. If choice is totally unpredictable, then any further empirical analysis will provide little if any significant results.

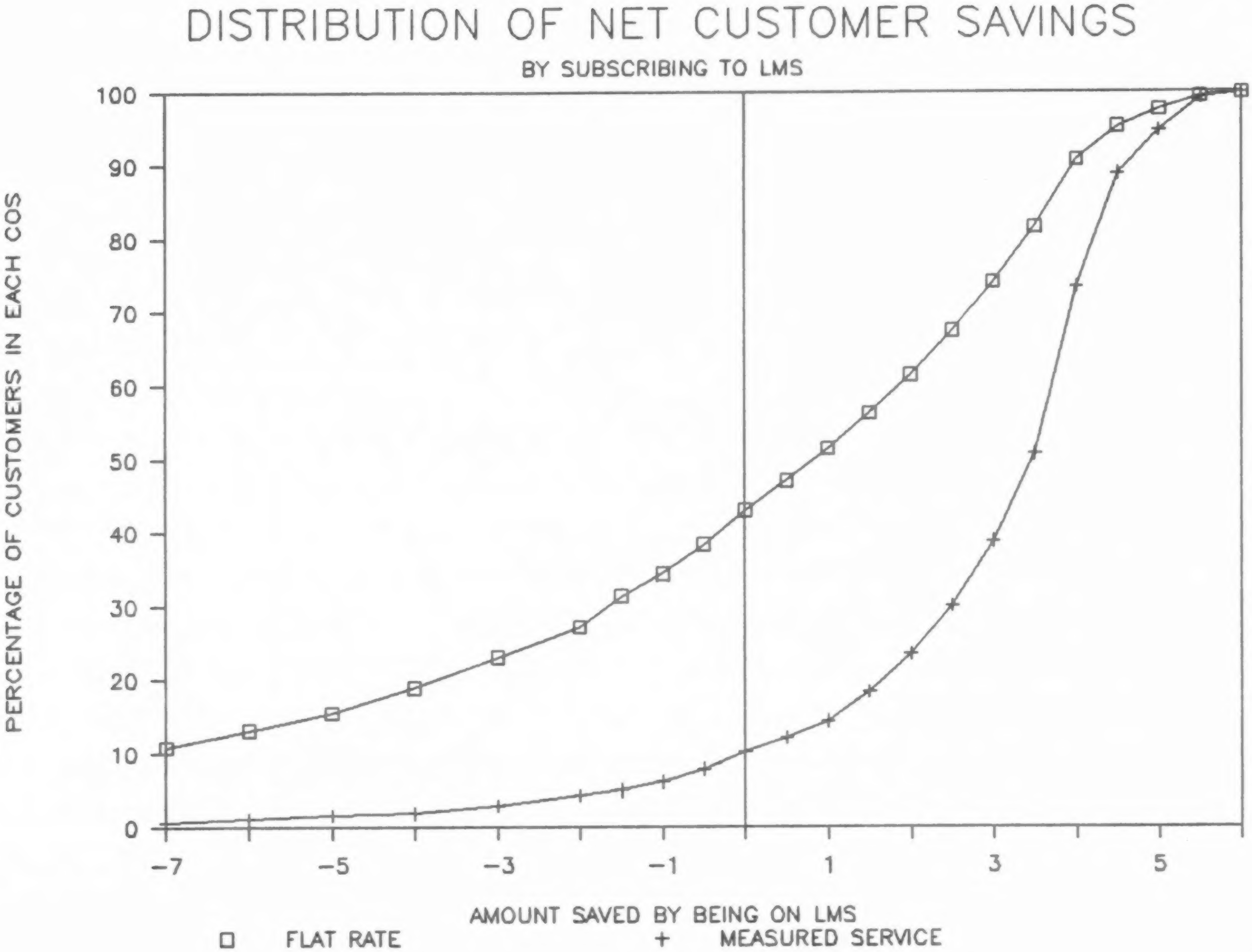
In order to undertake a preliminary examination of this question, 8 months of data on household calling behavior were studied (February through September, 1985). For this data set, only households that did not alter their class of service during those ten months were studied. This selection criterion was necessary since some customers who changed from measured service to flat rate service were not properly monitored after they had changed their class of service. For each customer and for each ZUM–TOD rate the mean number of calls, variance of number of calling and durations and the skewness of calling and duration were calculated. The means of number and duration of local and ZUM calls were utilized to calculate the resulting bill under either flat–rate or LMS calling. We then calculated the saving from being on flat rate service. Let CF be the size of the telephone bill under flat rate service and CM be the equivalent figure under LMS. Subtracting the RHS terms in equation 1 from those in equation 2 gives a cost saving from being on measured service as

$$\begin{aligned} CF - CM &= P^{m}_{acc} - P^{f}_{acc} - \sum_{t=1}^{3} [\,(p_{o,t1} - p_{1,t1})\cdot \bar{y}_{t1} + p_{1,t1}\cdot \bar{x}_{t1}] \\ &\quad + \text{Min}\{ \sum_{t=1}^{3} \sum_{z=1}^{3} [\,(p_{o,tz} - p_{1,tz})\cdot \bar{y}_{tz} + p_{1,tz}\cdot \bar{x}_{tz}]\,,\ C_m \} \\ &= P^{m}_{acc} - P^{f}_{acc} - \overline{LOCREV} + CREDIT. \end{aligned}$$

A bar over a variable indicates an average over the ten months for this particular study. Figure 10 shows the distribution of average actual or potential savings from being on measured service. The vertical axis indicates the proportion of observed subscribers in a particular rate class whose savings on measured service are at least as small as the corresponding figure on the horizontal axis. Positive values along the horizontal axis indicate a household that would have been cost–minimizing better off by being on measured service. The line that is everywhere furthest to the left is the distribution of flat rate customers. The difference between the distributions is apparently large and the difference between their means is significant at the .0001 probability level using a two–tailed t–test with different underlying population variances. The number of flat rate customers who are actually (cost–minimizing) better off under flat rate service is 43.0 per cent. On the other hand only 10.1 per cent of measured service customers would have been better off under flat rate service.

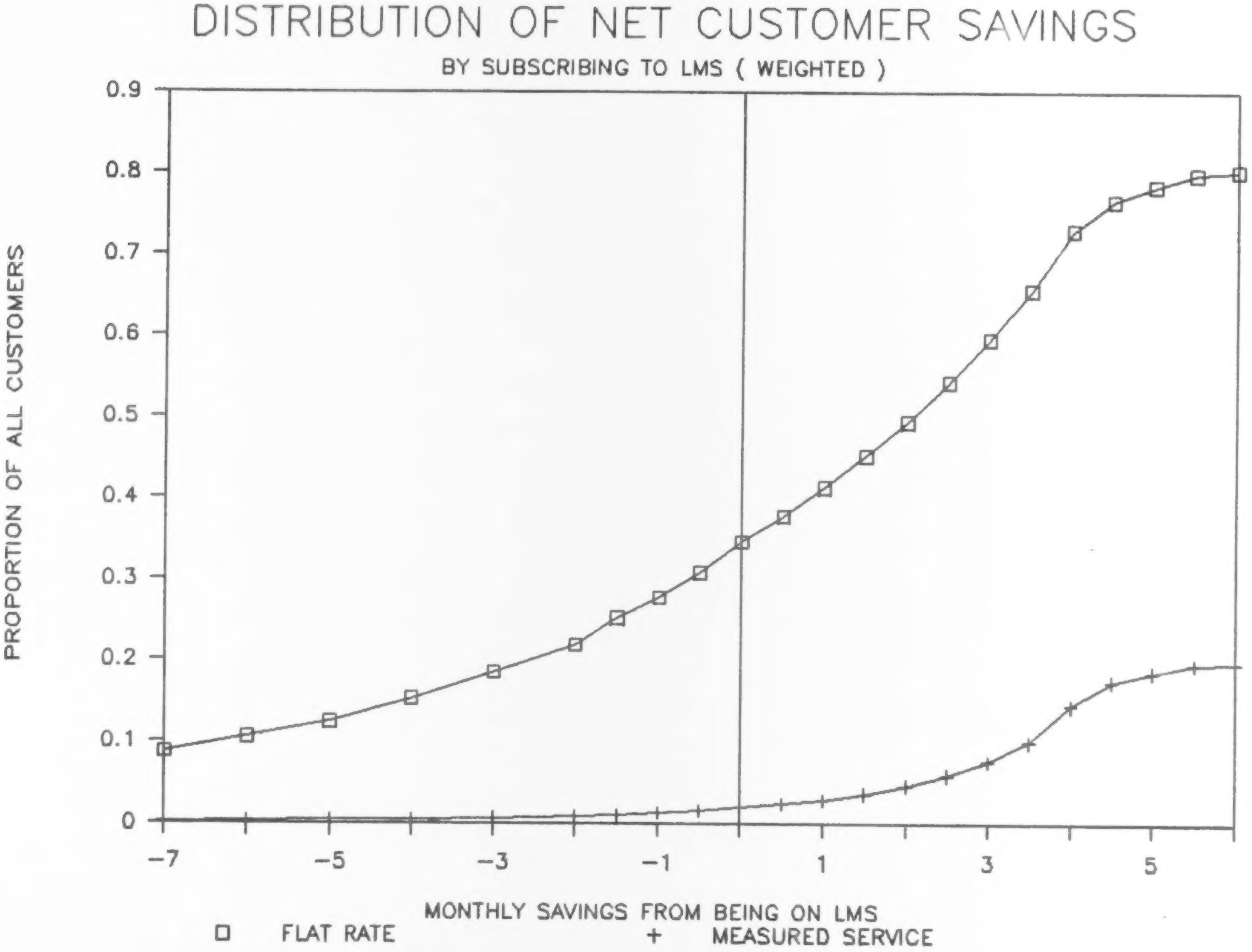
Figure ~~9~~ 10 displays the distribution functions to reflect the proportions of persons on measured and flat service.

Those on measured service are thus far more likely to have made the correct decision. This is largely due to the fact that most customers would be better off under


DISTRIBUTION OF NET CUSTOMER SAVINGS
BY SUBSCRIBING TO LMS
PERCENTAGE OF CUSTOMERS IN EACH COS
100
90
80
70
60
50
40
30
20
10
0
−7
−5
−3
−1
1
3
5
AMOUNT SAVED BY BEING ON LMS
□ FLAT RATE
+ MEASURED SERVICE

P. 28 (c)

FIG. 10



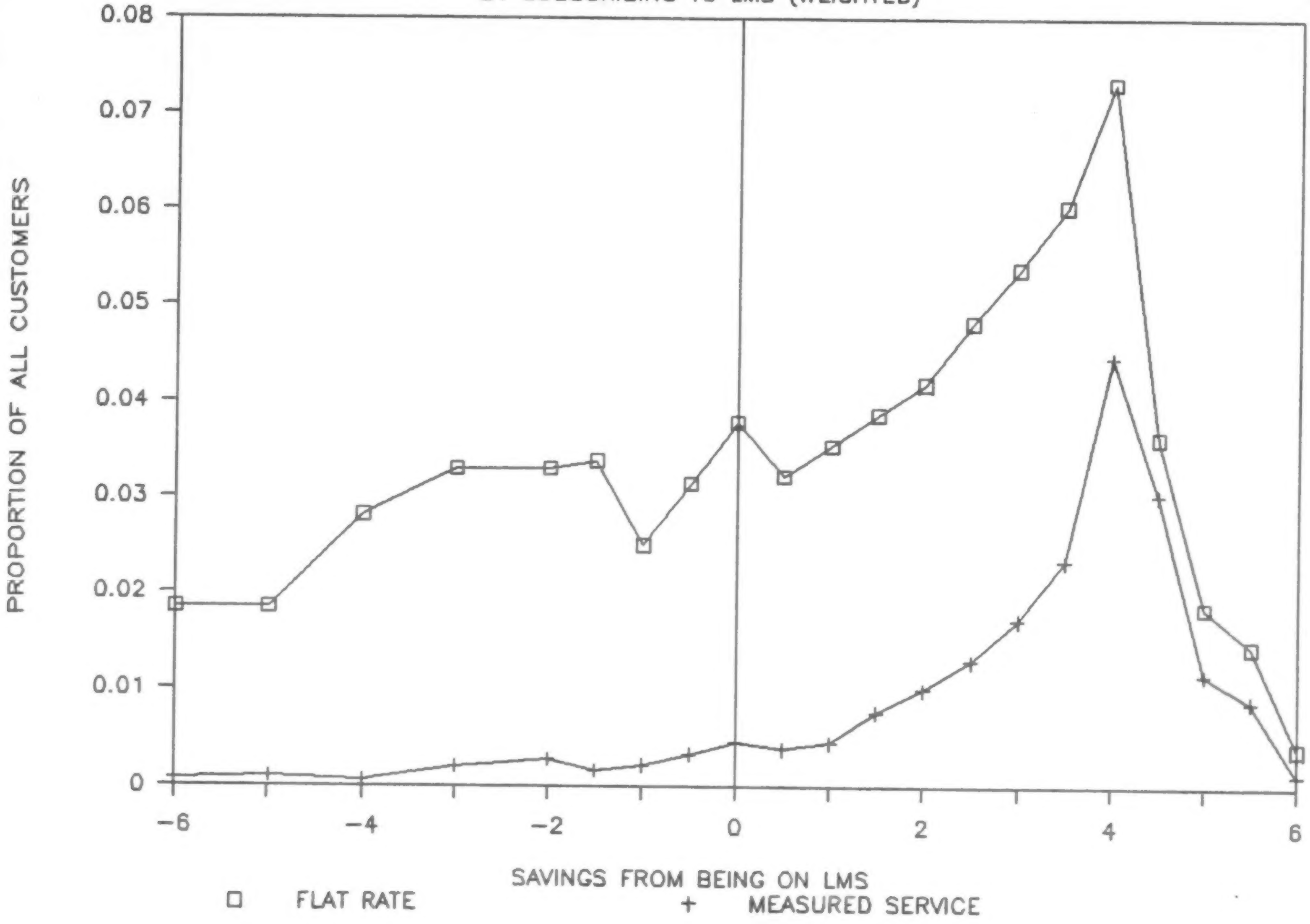
P. 28(b)

FIG. 11(a)

11 (a)

FREQUENCY DISTRIBUTION OF NET SAVINGS

BY SUBSCRIBING TO LMS (WEIGHTED)



P. 28(c)

Fig. 11(b)

measured service. On the other hand, there is also a considerable difference in the cost of a mistake when one compares the cost of making too few calls when one is on flat rate service and too many when one is on LMS. If one is on flat rate service, the cost of the making a mistake of underutilization (i.e. being to the left of $\bar{N}$ in Figure 9,) is, at the most, $6.40. Being on the right side of $\bar{N}$ while on measured service, on the other hand, can result in very high local calling bills. It seems that a great many telephone users consider $3.80 to be a reasonable price to pay for insurance against higher–than–expected telephone bills.

The maximum possible loss is greater than merely the difference in the cost of access, as can be seen by examining the equation for CF – CM. If a subscriber is on flat rate service, but makes only ZUM calls, then that customer is not only paying a higher access fee but is also foregoing $3.00 in credits that can be applied against ZUM calling. 12.0 per cent of customers on flat rate service actually have a average monthly loss of over $3.80.

Table 3 indicates the average standard deviations of the number of calls and of the total amount time spent on the telephone for calls placed into the local calling area and into zones 2 and 3. All the customers included in this table would have had lower bills had they been on measured service. The first column of results are for those customers who were on flat rate service, while the second column is for those customers who made the "correct", cost–minimizing to be on measured service. For every customer who fitted the criteria described above, standard deviations were computed for calls placed over the eight month period and then averages were taken for all customers. In all cases but one the variability of calling was greatest for flat rate customers.

Table 3

Means of Standard Deviations of Monthly Calling
Of Subscribers for Whom Measured Service
Is the Least Costly Telephone Service

			Standard Deviation	
			Flat Rate Customers	LMS Customers
Number of Calls				
	Local	Day	20.02	14.82
		Evening	13.72	10.12
		Night	15.05	11.15
	Zone 2	Day	2.72	2.28
		Evening	1.54	1.55
		Night	1.87	1.68
	Zone 3	Day	2.29	1.76
		Evening	1.48	1.24
		Night	1.52	1.32
Duration of Calls				
	Local	Day	81.32	57.69
		Evening	78.01	60.46
		Night	70.50	53.19
	Zone 2	Day	9.52	7.09
		Evening	11.90	12.58
		Night	10.55	15.71
	Zone 3	Day	9.71	7.52
		Evening	9.99	9.76
		Night	9.00	7.59

Number of observations = 4,004.

The differences in the means is significant at greater than 99 per cent for all local calling figures (i.e. number and duration of calls) but are insignificant for all the other types of calls or durations.[3] Flat rate customers tend to have greater variability in their calling patterns, thus increasing the probability that they will have unexpectedly large bills from local calling and inducing the flat rate class of service choice. It also makes it more difficult to make an accurate estimate of the expected amount of calling.

Other statistical analysis undertaken to date but not presented here strengthens the contention that people do monitor their calling behavior and that they make class of service choices and adjustments accordingly. An analysis of persons who changed the class of service indicate that, for instance, those persons who switch from flat to measured service made significantly fewer local calls than those who switched the other way. The number of toll calls was essentially the same. Furthermore, switching from flat rate service to measured service does, on average, result in reduced calling. The converse is also true.

---

[3] The significance of the difference was established using a t test to compare the means. This was done for both the the means of the natural logs of the standard deviation as well as the untransformed standard deviations. Logs were taken because the untransformed data was heavily skewed to the right rendering the t test inapprporiate. However, taking logarithms did not alter the result on the differences of the means..

The results of examining the ten months of data indicate, first, that potential subscribers who use the telephone more for making local and ZUM calls are more likely to subscribe to flat rate service. Considering the small cost of making an error on the side of caution, it is encouraging that this tendency is measurable at all. Secondly, those flat–rate subscribers who should, in a cost–minimizing sense, subscribe to measured service tend to have greater variability in their calling patterns than do measured service customers who have made the cost–minimizing class–of–service choice.

Wenders (1987, p. 127) suggests other explanations of the preference for flat rate service. First, a household will have to compare the marginal benefit of making a call with the cost of that call. Under measured service, this will require knowing the price structure, including the time of day that prices change. Under flat rate service such considerations are not necessary, thus eliminating "demand–side" transactions costs. Since these transactions costs reduce the marginal benefit of a call a household can increase the value of calling by simply choosing flat–rate service. In addition, Wenders claims that flat rate increase the value of incoming calls. This is presumably because incoming calls induce a desire to make more calls, and the more calls that one can make in response to incoming calls the greater the value of those calls. Another reason that flat rate service might be more valuable is that the value of the option to make more calls will also be greater.

Wenders reasons for preferring flat rate service introduce possible problems of endogeneity into the class of service choice exercise. If a subscriber wishes to enhance the option value of his telephone then he will purchase flat rate service and, presumably, exercise the option more often than if he had taken measured service.

6)A Model of Utility Maximization of Telephone Usage

The discussion in the previous section was based upon the direct utility functions. While use of direct (rather than indirect) utility functions present considerable problems in deriving estimatable demand functions (Deaton (1978) and Gorman (1976)) they do have the advantage of being intuitively understandable. Since a good deal of work must be undertaken in comparing the value of the utility function at different possible solutions and in finding the *maximum maxorum* of several points that may satisfy tangency conditions it is necessary to be able to derive demand functions from the direct utility function.

The approach here is to develop equations that describe the production of "specific satisfaction" (using Gorman's terminology) from the use of the telephone and incorporate that into a utility function. From that we try to develop demand functions.

To fix ideas: Suppose that the vector of inputs of telephone related goods is denoted t and that is separable from all other goods consumed, C. That is, the household budgets itself to spend a certain amount on telephoning and then allocates that amount on calls during different time of day periods and zones. Then the direct utility function can be written as

$$U = f(t,C) = F(\tau(t),C)$$

$\tau(t)$ can be considered to be a production function yielding "specific satisfaction" (in the telephone case, information or entertainment) which we will call Z. (i.e. $Z = \tau(t)$).

Let $t = (N_{tz}, D_{tz}, \psi)$ where N_{tz} is the number of completed calls made during period t to zone z and D_{tz} is the average duration of those calls. ψ indicates a vector of the values of household characteristics and of all other inputs to the production of information/entertainment. These inputs will include the degree of access (e.g. the number of other telephone subscribers the local calling area) and the total amount of time devoted to telecommunications. The index t indicates the time of day period (t = day, evening, night), z indicates the zones being called (z = zone 1, zone 2, zone 3 or intra–LATA toll).

Any production function of specific satisfaction must have certain properties that satisfy both technical considerations and intuitive priors about how telephone customers behave. The first of these is that, as time spent on the telephone goes to zero, the utility of calling also goes to zero since the information content of short calls will be low and the transactions costs are presumably significant. Expressed another way, the condition states that for the same number of calls, utility is greater the longer is the average duration, at least over some range of possible durations. Such a property of the utility function will eliminate the intuitively unappealing possibility that telephone customers would achieve positive utility of telephone use by making no calls that last for a total of some positive number of minutes. It also eliminates the equally unappealing possibility that an infinite number of calls and very short total amount of time spent on the telephone represents the utility maximizing bundle of telephone goods.

The law of diminishing returns should also be satisfied: the marginal product of all calling should be positive but decreasing. In fact, utility should decrease once total duration or total calling have reached a certain level. This property indicates the greater opportunity cost of time as total time spent on the telephone increases

The production function should also accommodate the large number of households that make no calls to some zones during some time-of-day periods. That is, the production function should be capable of achieving a maximum very close to zero calls for some values of the parameters.

Let x_{tz} be the total duration of all calls made to zone z during time-of-day period t and let y_{tz} be the number of calls made to z during t. We assume that there are two functions, $g_{tz}(x_{tz})$ and $h_{tz}(y_{tz})$, and that these functions are multiplied together to produce utility for each TOD-zone combination. That is $u_{tz} = f(\, g_{tz}(x_{tz}) \cdot h(y_{tz})\,)$. The properties outlined above are:

$$\lim_{\substack{x\, =\, K \\ y\, \to\, \infty \\ x/y\, \to\, 0}} f(x_{tz}, y_{tz}) = 0,$$

$$f_{x_{tz}} > 0 \quad f_{x_{tz}x_{tz}} < 0$$

$$f_{y_{tz}} > 0 \quad f_{y_{tz}y_{tz}} < 0$$

$$f_{y_{tz}x_{tz}} < 0$$

where K is some constant. These conditions can be assured if $h_{tz}(y_{tz}) \to 0$ as $y_{tz} \to \infty$ or as $y_{tz} \to 0$ by virtue of the fact that the two functions enter $f(\cdot)$ multiplicatively. For the two subfunctions we propose the forms:

$$g_{tz}(x_{tz}) = x_{tz}^{\gamma}(\theta_{tz} - x_{tz}) \text{ and}$$
$$h_{tz}(y_{tz}) = y_{tz}^{\beta}(\alpha_{tz} - y_{tz}).$$

In addition the function $f(\cdot)$ is multiplied by time and zone specific parameters τ_s and ρ_w where s and w refer to times and locations not necessarily equal to t and z. This will allow the utilization of the model to test the impact of different tariff structures. For now we set $s = t$ and $w = z$. The parameters can themselves be functions of the household's characteristics. The utility function is defined over all goods, y_{tz}, x_{tz} and a

Hicksian composite commodity, C, made up of all non–telephone goods:

$$U = \delta \ln C + \Sigma_t \Sigma_z \tau_t \rho_z x^{\gamma}_{tz}(\theta_{tz} - x_{tz}) y^{\beta}_{tz}(\alpha_{tz} - y_{tz})$$

which is maximized subject to the budget constraint

$$\Sigma_t \Sigma_z (P^y_{tz} y_{tz} + P^x_{tz} x_{tz}) + C + P_{acc}$$

where P^y_{tz} is the net cost of placing a call to z during t, P^x_{tz} is the cost of each minute of calling and P_{acc} is the cost of access to the telephone system.

The earning of income may, of course, not be exogenous to calling. Many telephone users work out of their home and many persons receive job offers over the telephone at their place of residence. For most persons use of a telephone must make some impact on their full income. At present, this model takes no account of this endogeneity problem and may therefore provide biased results.

The Lagrangian of this problem is

$$\Lambda = U = \delta \ln C + \Sigma_t \Sigma_z \tau_t \rho_z x^{\gamma}_{tz}(\theta_{tz} - x_{tz}) y^{\beta}_{tz}(\alpha_{tz} - y_{tz})$$

$$\lambda \{\Sigma_t \Sigma_z P^y_{tz} y_{tz} + \Sigma_t \Sigma_z P^x_{tz} x_{tz} + C + P_{acc}\}$$

Dropping the τ and ρ parameters, the resulting first order conditions are:

i) $$\delta/_C = \lambda$$

ii) $$x^{\gamma-1}_{tz} y^{\beta}_{tz}(\alpha_{tz} - y_{tz})(\gamma\theta_{tz} - (\gamma + 1)x_{tz}) = \lambda P^x_{tz}$$

iii) $$x^{\gamma}_{tz} y^{\beta-1}_{tz}(\theta_{tz} - x_{tz})(\beta\alpha_{tz} - (\beta + 1)y_{tz}) = \lambda P^y_{tz}$$

and the budget constraint. Note that conditions ii) and iii) can be solved for total duration of calls for each time–ZUM period when prices are zero.

$$y_{tz} = {}^{\alpha_{tz}\beta}/_{(\beta+1)} \text{ when } P^y_{tz} = 0$$

and

$$x_{tz} = {}^{\theta_{tz}\gamma}/_{(\gamma+1)} \text{ when } P^x_{tz} = 0.$$

Equations ii) and iii) can be solved to give us a function that is quadratic in both duration, x_{tz}, and in number of calls, y_{tz}. The solution for x_{tz} is:

iv) $$x_{tz} = \frac{-B_{tz} - \sqrt{B^2_{tz} - 4\ A_{tz} D_{tz} y_{tz}}}{2A_{tz}}$$

where $$A_{tz} = (\beta\alpha_{tz} - (\beta+1)y_{tz})P^x_{tz}$$

$$B_{tz} = -[\ \theta_{tz}(\beta\alpha_{tz} - (\beta+1)y_{tz})P^x_{tz} + y_{tz}(\alpha_{tz} - y_{tz})(\gamma + 1)P^y_{tz}\ ]$$

$$D_{tz} = (\alpha_{tz} - y_{tz})\gamma\theta_{tz}P^y_{tz}$$

The sign before the discriminant must be negative. If it is not then $x_{tz} = \theta_{tz}$ when $y_{tz} = 0$ clearly a violation of the condition that total duration goes to zero as the number of calls goes to zero. Solving i) and iii) for C and substituting iv) into the resulting equation gives us

v) $$C = \delta P^y_{tz} y^{1-\beta} \left[\frac{2A}{-B - \sqrt{B^2 - 4ADy_{tz}}} \right]^{\gamma} \left(\theta_{tz} - \left[\frac{-B - \sqrt{B^2 - 4\ ADy_{tz}}}{2A} \right] \right)^{-1} (\beta\alpha_{tz} - (\beta+1)y_{tz})^{-1}$$

where, for convenience the subscripts have been dropped from A, B and D.

This, unfortunately, is as far as this utility function can be taken in the derivation of a demand function of the form $x_{tz} = l(P^x_{tz}, w_t, M)$. Since the variables cannot be solved in closed form as functions of each other and of prices, they cannot be substituted into the budget equation to find demand functions.

However, the values of C and of the x_{tz}s and y_{tz}s can be solved by numerical methods. Since equation v) has at least four roots it would be helpful if restrictions could be placed over the range of the equation. Fortunately, a restriction exists that limit the number of roots over the range of interest to one. These restrictions arise out of economic considerations and can best be discussed by referring to the general shape of the equation, shown in Figure 11. (This is the form as long as the restriction on the sign

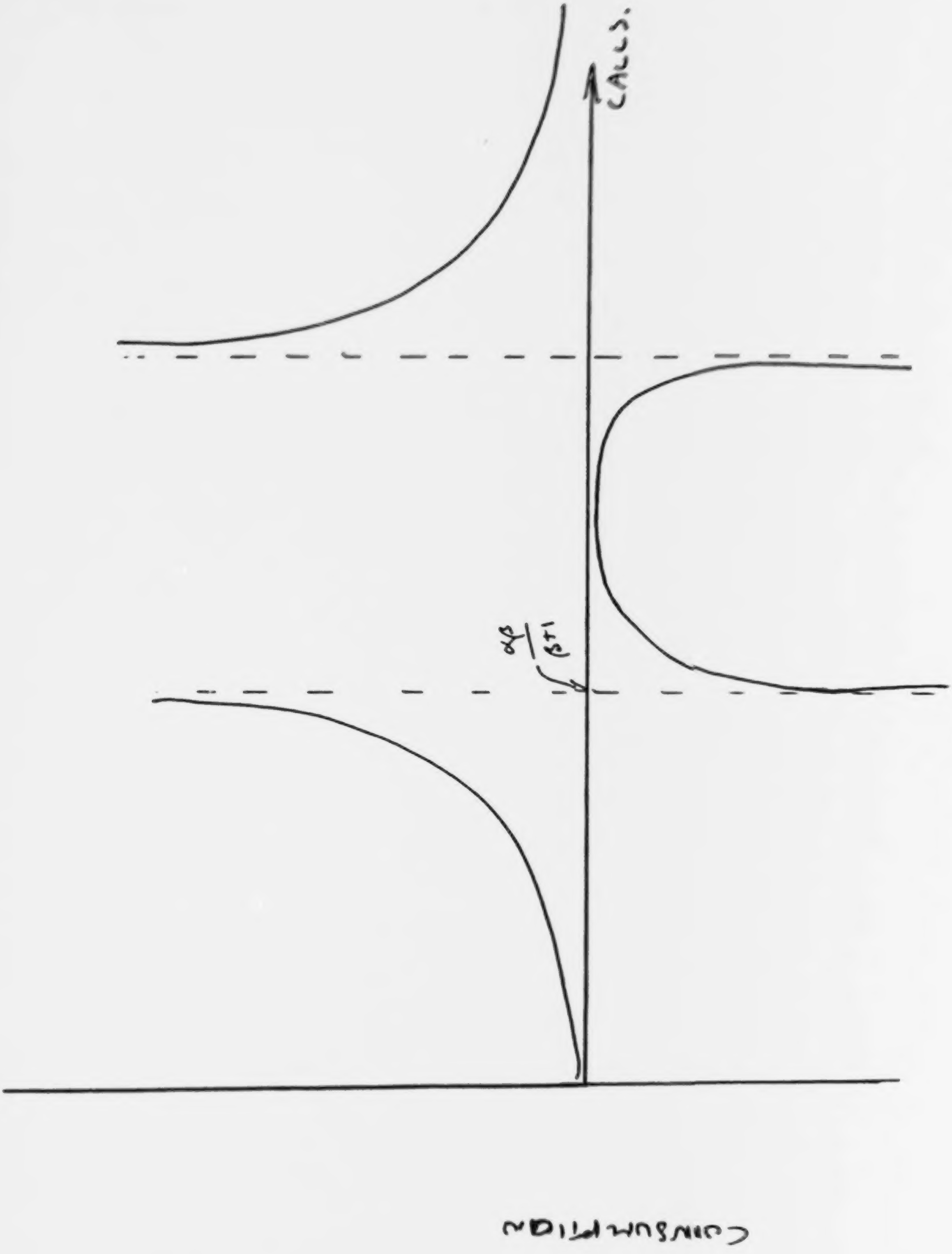


FIG. 11

of the discriminant in the equation for x_{tz} is maintained.) Since Figure 11 is displaying an income expansion path we can rule out the portion of the function below the horizontal axis because negative consumption is not allowed. Similarly we are assuming that both consumption and calling are not inferior goods. This eliminates the portion of the function that is furthest from the origin.

A program to solve such a system has been written for customers on measured service with four zones to call and three time–of–day pricing periods. The technique is to search over values of C, find a set of calls and durations that maximize utility given the experimental value of z and then calculate the extent to which the budget constraint is violated. Using Brent algorithms (from the IMSL Fortran Subroutine Library) this process is repeated until the roots of the first–order condition in v) and the budget constraint are found. As a (purely hypothetical) demonstration, the results of simulation for a hypothetical customer facing a synthetic pricing regime is presented on Table 4.

Table 4
Results of Synthetic Simulation

Assumptions: Income = 1250, $\alpha_{tz} = 800$, $\theta_{tz} = 300$, $\forall$ t,z
$\beta = .6$, $\gamma = .4$, $\tau^{*}\rho = .0001$.

Prices:
Calls py(TOD period,Zone)

py(1,1)	py(1,2)	py(1,3)	py(1,4)	py(2,1)	py(2,2)	py(2,3)	py(2,4)
.2	.4	.5	2.5	.12	.24	.3	1.5

py(3,1)	py(3,2)	py(3,3)	py(3,4)
.8	1.6	.2	1.0

Durations px(TOD period, Zone

px(1,1)	px(1,2)	px(1,3)	px(1,4)	px(2,1)	px(2,2)	px(2,3)	px(2,4)
.05	.15	.25	.5	.03	.09	.15	.3

px(3,1)	px(3,2)	px(3,3)	px(3,4)
.02	.06	.1	.2

Results: Non–telephone Consumption = 214.7

TOD period	Zone	Number of Calls	Total Time On Calls
Day	1	257.8	84.1
Day	2	220.6	81.0
Day	3	203.7	77.9
Day	4	40.3	58.4
Evening	1	274.0	84.8
Evening	2	249.9	82.9
Evening	3	238.6	81.1
Evening	4	90.5	73.3
Night	1	282.5	85.1
Night	2	265.8	83.8
Night	3	257.8	82.6
Night	4	136.2	78.5

Figure 12 shows indifference curves that are generated from solving equation v) for given levels of utility. The trade–off shown there is between number of calls on the horizontal axis and all–non telephone goods on the horizontal axis. Figure 13 shows the resulting demand functions that arise out of solving v).

The class of service and utilization decision is simulated in the following step–wise procedure. First, the customer is treated as choosing not amongst different segments of a non–linear, non–convex budget set but amongst several linear budget sets formed by extending each linear segment to the axes of the positive orthont. Figure 14 shows this with the dashed segments indicating the artificial extensions of the budget set for the tree possible classes of service. Of course, only one time–ZUM zone is shown; the actual extensions are over the 12 possible types of time–ZUM calls.

Next, for each linear budget set and for a given set of parameters, the number of calls and the total duration of calls in each time–ZUM zone are calculated from equation v). The level of utility is also calculated. Figure 15, where we consider only measured service calling in one time period and one zone, shows some possible results. The downward sloping section of the budget set results in a an optimum at A while the

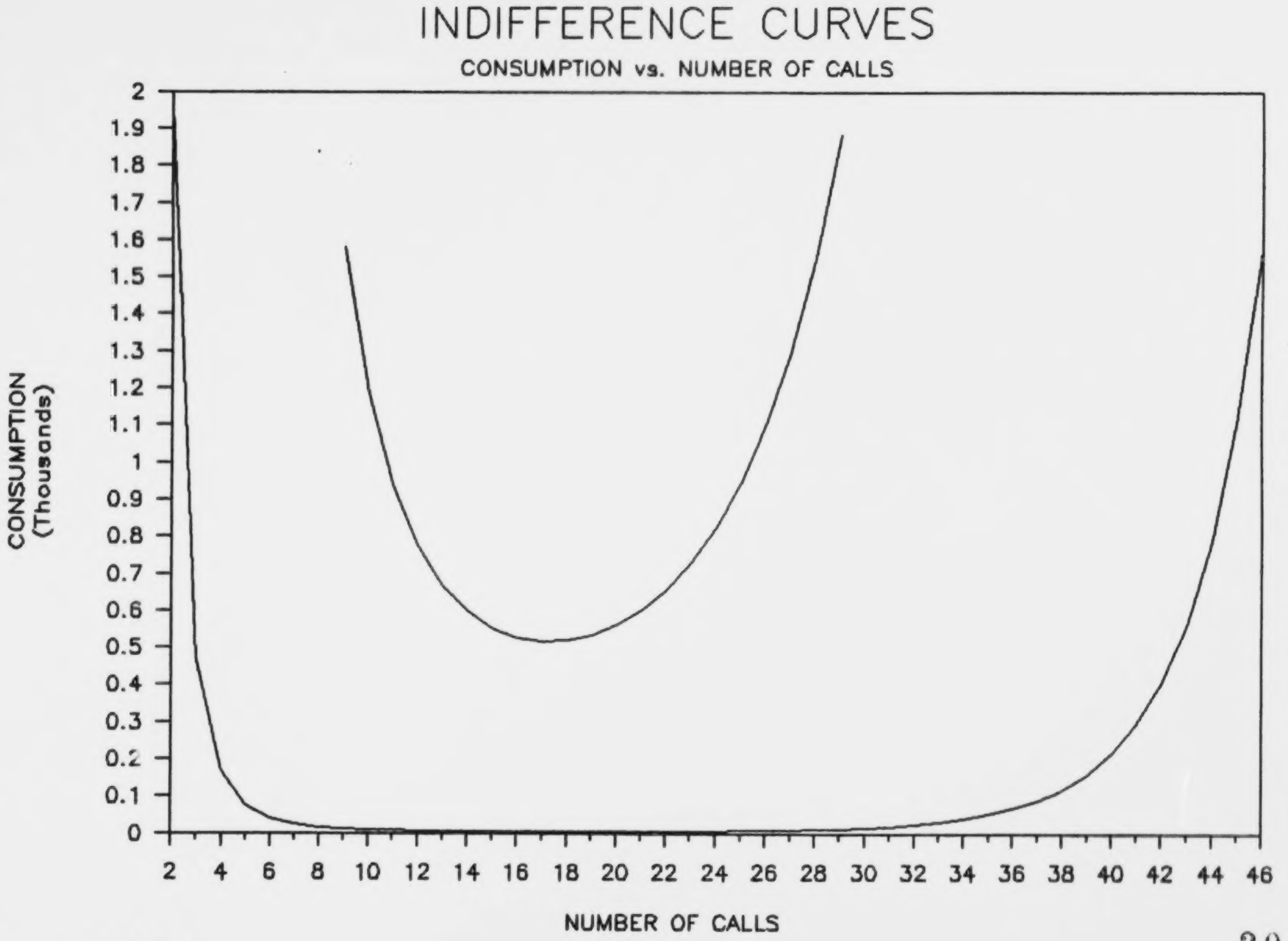

INDIFFERENCE CURVES
CONSUMPTION vs. NUMBER OF CALLS
CONSUMPTION (Thousands)
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
NUMBER OF CALLS

FIG. 12

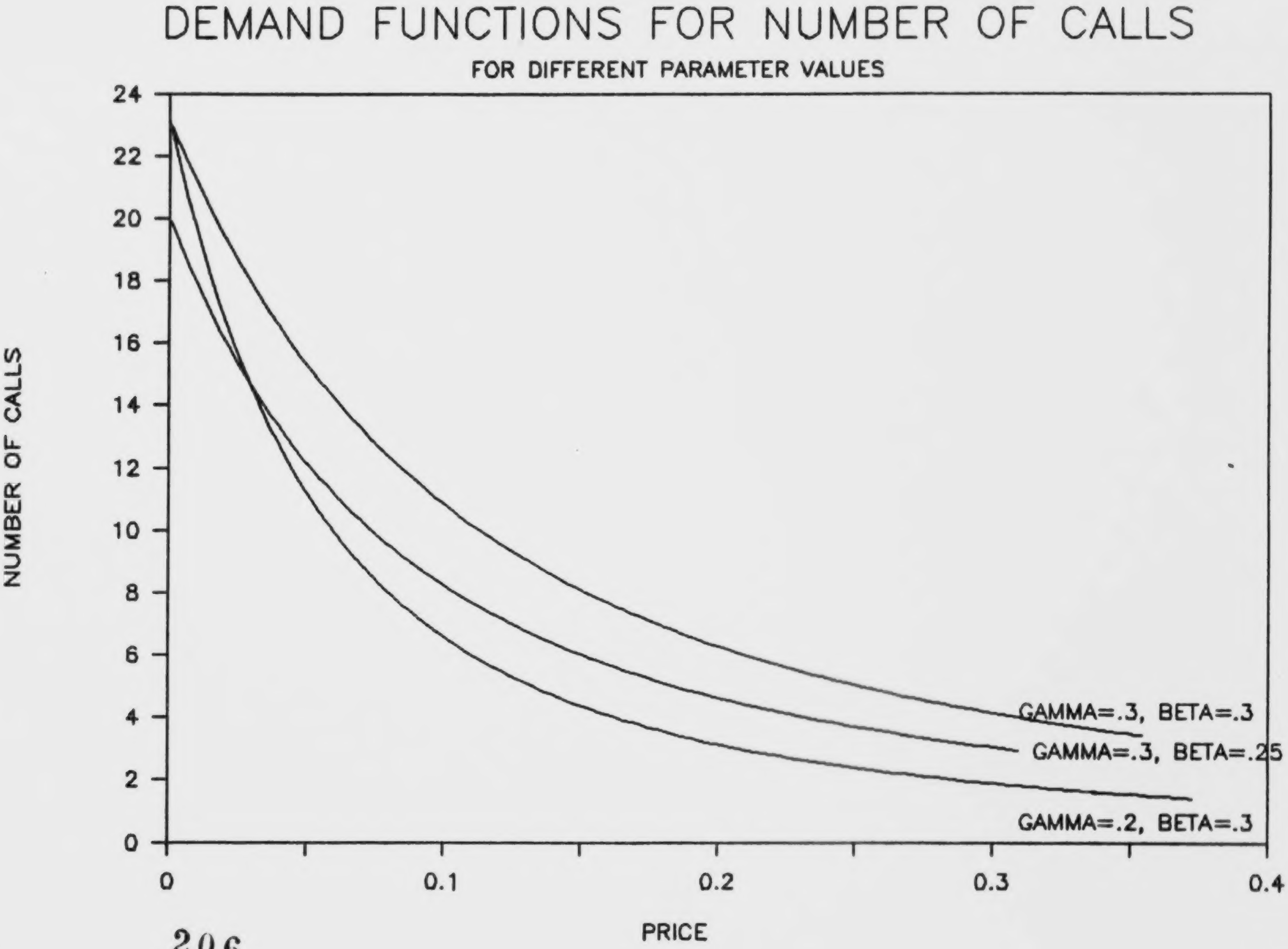
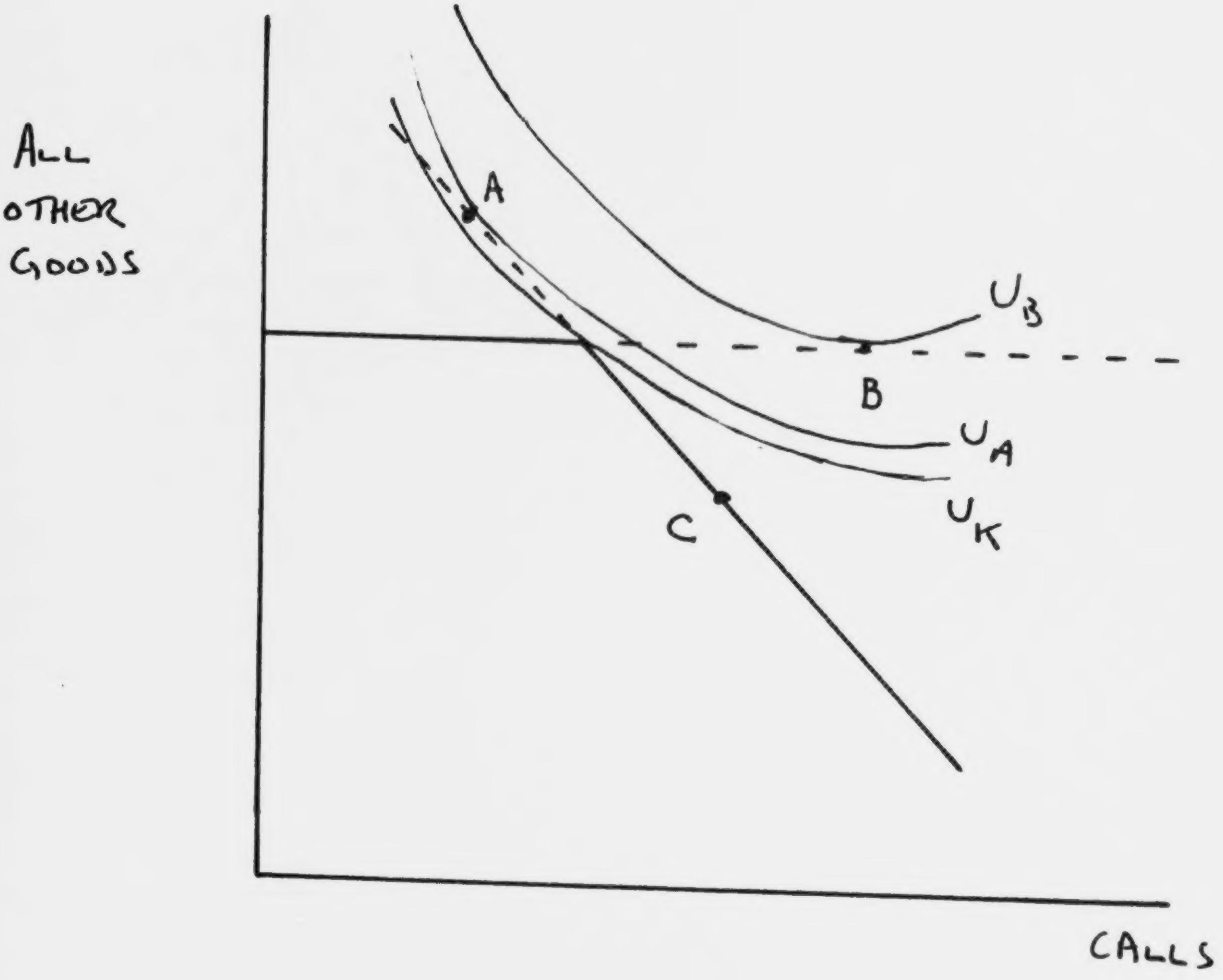

DEMAND FUNCTIONS FOR NUMBER OF CALLS
FOR DIFFERENT PARAMETER VALUES
NUMBER OF CALLS
24
22
20
18
16
14
12
10
8
6
4
2
0
0
0.1
0.2
0.3
0.4
PRICE
GAMMA=.3, BETA=.3
GAMMA=.3, BETA=.25
GAMMA=.2, BETA=.3

FIG. 13



FINDING THE UTILITY MAXIMIZING CHOICE OF CALLING WITH LMS

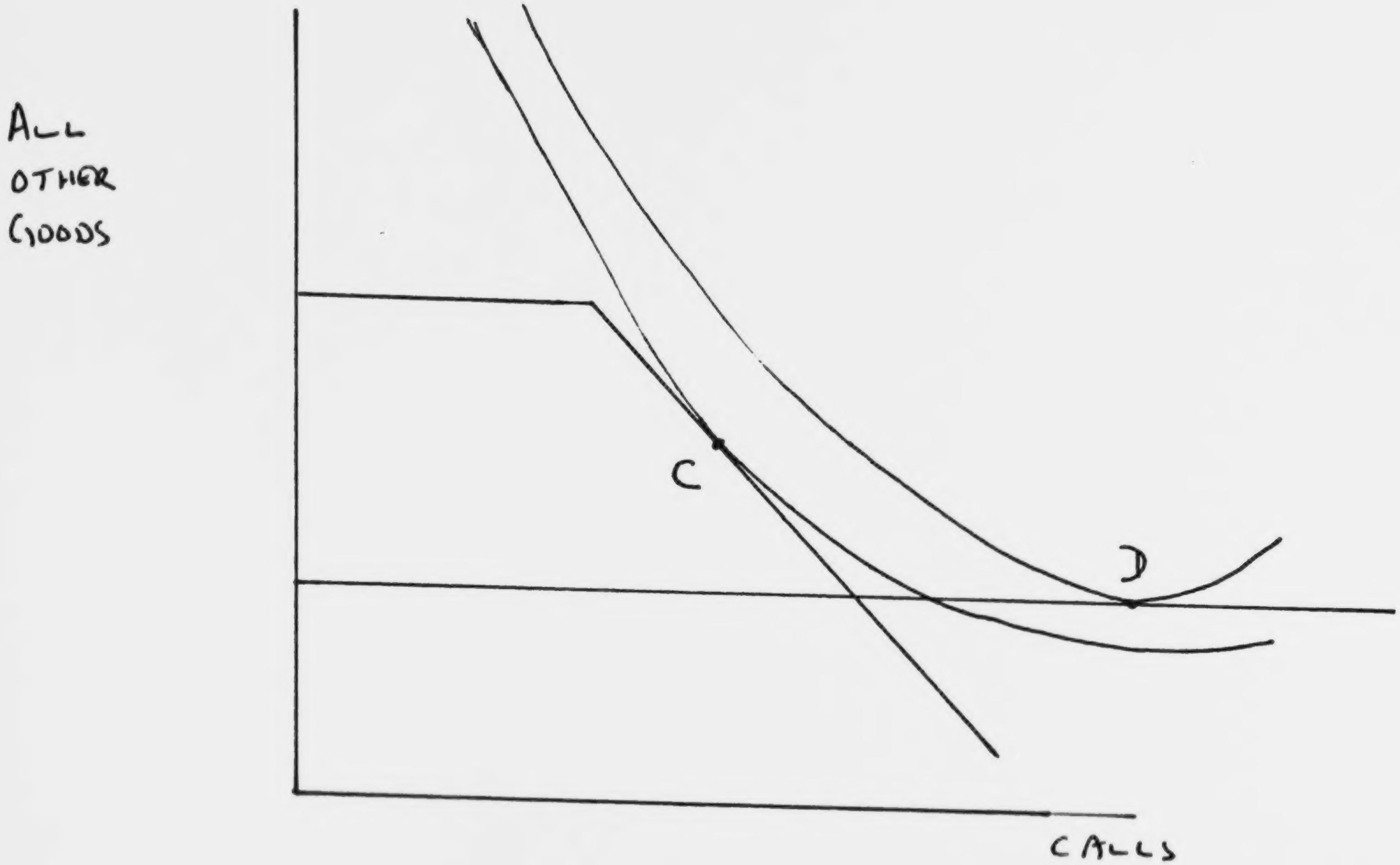
FIG. 15.

horizontal portion (the extension of the effect of the measured service credit) has an optimum at B. That indicates that the respective indifference curves lie approximately as indicated by U_A and U_B. Since neither A or B are possible the actual best choice when confronted with only measured service is the kink point, with the indifference curve looking like U_K.

If, however, C had been observed instead of A then C would be the only possible solution under measured service. In Figure 16 we see that comparing the measured service option with the budget set implied by flat rate service, point D might result in a higher level of utility.

7) Conclusions

This paper has presented a broad discussion of some of the technical and public policy issues related to telephone pricing. The work that is being proposed here is part of a general effort to increase the understanding of the demand for telephone services and of the costs of providing those services. The need for this sort of information has increased significantly since the Modified Final Judgment upset the relatively quiet life of telephone regulation and the previous structure of cross–subsidies. For regulators this information is necessary for effective regulation while for the Bell Operating Companies a complete understanding of the nature of demand will be necessary to make the strategic decisions that will reduce the impact of uneconomic bypass. In addition, the work is of interest in itself, dealing with difficult issues of estimation in the face of non–linear and non–convex budget sets, risk aversion and the recursive nature of the class of service choice.



COMPARING FLAT RATE CALLING UTILITY WITH MEASURED RATE CALLING UTILITY.

FIG. 16

3"ize)

References

Adams, W. J. and J. L. Yellen (1976) "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, pp. 475 – 498.

Artle, R, and C. Averous (1973), "The Telephone System as a Public Good: Static and Dynamic Aspects," *Bell Journal of Economics*, Vol. 4, pp. 89–100.

Alleman, J. H. (1977) "The Pricing of Local Telephone Service", U.S. Dept. of Commerce, Office of Telecommunications.

Baumol, W. J. and D. F. Bradford, (1970) "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, Vol. 60., pp 265 – 283.

Brander J.A. and B.J. Spencer (1983) "Local Telephone Pricing: Two–part Tariffs and Price Discrimination," in Leon Courville, Alain De Fontenay and Rodney Dobell, eds., "Economic Analysis of Telecommunications" North–Holland, New York, 1983. pp.305–316.

Braeutigam, Ronald R. (1979) "Optimal Pricing with Intermodal Competition" *American Economic Review*, 69(1) pp.38–49.

Brown, S. J. and D. S. Sibley A. "The Theory of Public Utility Pricing" Cambridge University Press, New York, 1986.

Burtless, G. and J. A. Hauseman, (1978) "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment", Journal of Political Economy, Vol. 86, no. 6, pp. 1103– 1130.

California Public Utility Commission (1985) "Charting a sustainable Regulatory Course in Telecommunications". Policy and Planning Division, San Francisco.

Coase, R. H. (1946) "The Marginal Cost Controversy," *Economica*, Vol. 13, pp. 169 – 189.

Cornell n. W. and R. G. Noll (1985) "Local Telephone Prices and the Telephone Question," paper presented at the Bell Communications Research Demand Analysis Forum, New Orleans, Oct.

Curien N. and E. Vilmin (1983) in Courville, Leon, Alain De Fontenay and Rodney Dobell, eds., *Economic Analysis of Telecommunications* North–Holland, New York, 1983. pp.231–249.

Deaton, Angus, (1978) "Specification and Testing in Applied Demand Analysis," *The Economic Journal*, Vol. 88, pp. 524–536, Sept.

________________ and John Muellbauer, (1980) *Economics and Consumer Behavior*, Cambridge University Press, New York.

Fischetti, M. A. "By–passing is Big Business for Business," *IEEE Spectrum*, Nov.

1985, pp. 78 – 83.

Goldman, M. B., H. E. Leland and D. S. Sibley (1984) "Optimal Non–uniform Pricing," *Review of Economic Studies*, April, 1984.

Gorman, W. M., (1976) "Tricks with Utility Functions," in M. J. Artis and A. R. Nobay (eds.) *Essays in Economic Analysis*, Cambridge University Press, New York

Graham, J. W. and C. A. Green (1984) "Estimating the Parameters of a Household Production Function with Joint Products," *Review of Economics and Statistics*, Vol. 66, pp. 277 – 283.

Griffen, James M. (1982) "The Welfare Implications of Externalities and Price Elasticities for Telecommunications Pricing," *The Review of Economics and Statistics*, Vol. 64, No. 1, pp. 59 – 66.

Gronau, R. (1977) "Leisure, Home Production, and Work – the Theory of the Allocation of Time Revisted," Journal of Political Economy, Vol. 85, No. 6, pp. 1099 – 1123.

Hansen, L. P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, No. 4 pp. 1029–1054, July.

Hauseman, J. A. (198x) "Labor Supply" in

Henderson, J. M. and R.E. Quandt (1980) *Microeconomic Theory* McGraw–Hill, New York.

Huber, Peter W., (1987) (1987) "The Geodesic Network: 1987 Report on Competition in the Telephone Industry," Report issued by the U.S. Dept. of Justice, Antitrust Division.

Infosino, W. J. (1980) "Relationships Between the Demand for Local Telephone Calls and Household Characteristics," *Bell System Technical Journal*, Vol. 59, No. 6, pp. 931–953.

Jackson, C. L. and J. H. Rohlfs (1984), "Access Charging and Bypass Adoption, A Study Prepared for Bell Atlantic," Shooshan and Jackson, Inc., Washington, D. C.

Kahn, A. E. (1983) Testimony 82–11–07, 83–01–21, 83–04–02 before the California Public Utilities Commission.

Kahn, A. E. (1984) "The Road to More Intelligent Telephone Pricing," *Yale Journal on Regulation*, Vol. 1, pp. 189 – 157.

Kahn, A. E. and W. B. Shew (1987) "Current Issues in Telecommunications: Pricing," *Yale Journal on Regulation*, Vol. 4, pp. 191 – 256.

Kooreman, P. and A. Kapteyn (1987) "A disaggregated Analysis of the Allocation of Time within the Household," *Journal of Political Economy*, Vol. 95, No. 2, pp. 223 – 249.

Leland, Hayne E. and Robert A. Meyer (1976) "Monopoly Pricing Structures with Imperfect Discrimination." *Bell Journal of Economics*. Vol 7, No. 2, Aug.

Mitchell, Bridger (1978) "Optimal Pricing of Local Telephone Service " *American Economic Review*, Vol. 68, No. 4 pp. 517–537, Sept.

Naughton, Michael C. (1986) "The Efficiency and Equity Consequences of Two–part Tariffs in Electricity Pricing," *The Review of Economics and Statistics*, Vol. 68, No. 3, pp. 406 – 414.

Ng, Y. and W. Weisser, (1974), "Optimal Pricing with a Budget Constraint – The Case of the Two–part Tariff," *Review of Economic Studies*, Vol. 41, pp. 337 –345, July.

Oi, Walter Y. (1971) " A Disneyland Dilemma: Two–part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics*, Vol 85, pp. 77–96 (Feb.)

Park, Rolla Edward and Bridger M. Mitchell (1986) "Optimal Peak–Load Pricing for Local Telephone Calls". and Technical Appendix. Rand Corporation,R–3404–RC

Park, R. E., B. M. Mitchell, B.M. Wetzel and J. H. Alleman (1983) "Charging for Local Telephone Calls, How Household Characteristics Affect the Distribution of Calls in the GTE Illinois Experiment," *Journal of Econometrics*, Vol 22, pp. 339–364

Park, R. E., B.M. Wetzel and B. M. Mitchell (1983) "Price Elasticities for Local Telephone Calls," *Econometrica*, Vol. 51, No. 6, pp. 1699–1730.

Perl, Lewis J. (1984) "Residential Demand for Telephone Service: Preliminary Results" in Mann, Patrick C. and Harry M. Trebing, eds. Changing Patterns in Regulation, Markets, and Technology: The Effect on Public Utility Pricing. MSU Public Utilities Papers, 1984

Perl, Lewis J. (1984) "Residential Demand for Telephone Service 1983" Prepared for the Central Services Organization, Inc (now Bell Communications Research) by National Economic Research Associates, Inc. White Plains, New York.

Phillips, A. and G. L. Roberts (1985) "Borrowing From Peter to Pay Paul: More on Departures of Price from Marginal Cost," in Antitrust and Regulation: Essays in Honor of John J. McGowen, F.M. Fisher (ed), MIT Press, Cambridge, Mass.

Rohlfs, J. (1979) "Economically Efficient Bell System Pricing," Bell Laboratories Economic Discussion Paper No. 138

Sandler, Todd and John T. Tschirhart (1980) "The Economic Theory of Clubs: An

Evaluative Survey," *Journal of Economic Literature*, Vol. 18, pp. 1481–1521, Dec..

Schmalensee, R., (1981) "Monopolistic Two–part Pricing Arrangements," *Bell Journal of Economics*, Vol. 12, pp. 445–466.

Taylor, L.D. (1980) "Telecommunications Demand: A Survey and Critique". Ballinger, Cambridge, Mass.

Taylor, L.D. and D. Kridel (1985) "Residential Demand for Access to the Telephone Network Using Census Tract Data" presented at the Telecommunications Demand Modeling Conference, New Orleans, 23–25 Oct., 1985

Telecommunications Reports (1985) "Bypass Study, Predicting Rapid Substitution of Special Access.." Vol 51, No. 15 pp. 16–17, April 15.

Telecommunications Reports (1986a) "Nebraska First State to Enact Removal of Basic Local Services from PSC Rate Base Regulation,..." Vol 52, No. 16 pp. 3–4, April 21.

Telecommunications Reports (1986b) "In Exchange for Regulatory Flexibility, Pacific Bell Offers Basic Rate Predictability," Vol 52, No. 20 p. 20, May 19.

Telecommunications Reports (1986c) "Battle over New Nebraska Telecommunications Law Moves to the Courts; Both Sides Sue," Vol 52, No. 22 p. 37, June 2.

Telecommunications Reports (1986d) "Alabama Approves Rate Stabilization Plan..." Vol 52, No. 45, pp. 3–4, Nov. 10.

Telecommunications Reports (1987a) "...Vermont Legislative Panel Backs 'Social Contract' Bill;..." Vol 53, No. 5 pp. 1–2, Feb. 2.

Telecommunications Reports (1987b) "Finding no Major Changes in Exchange Bottleneck or in FCC regulation, Greene says Core Restrictions Must Stand;..." and "Customer Bypass has not Ended Local Exchange Bottleneck..." Vol. 53, No. 37, pp.1–5 and pp. 5–6, Sept.14.

Train, K. E., D. L. McFadden and M. Ben–Akiva (1986) "The demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *Rand Journal of Economics.*

Varian, H. (1978) "Microeconomic Analysis" First Edition, Norton, New York.

Wales, T. J. and A. D.Woodland (1977) "Estimation of the Allocation of Time for Work, Leisure and Housework", *Econometrica*, Vol. 45, No. 1, pp. 115 – 132.

Willig, R.C. (1976) "Consumer's Surplus Without Apology" *American Economic Review*, Vol. 66, pp.589–97.

Willig, R.C. (1978) "Pareto–superior Non–linear Outlay Schedules," *Bell Journal of Economics*, Vol. 9, pp. 56–69, Spring.

Appendix A Explicit Statements of The Budget Set

The pricing of telephone service results in a complex set of choices for a potential telephone subscriber. The budget constraint for a household on measured service rate is

$$B = \sum_{t=1}^{3} \sum_{z=1}^{4} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}]$$

$$- \mathrm{Min}\{ \sum_{t=1}^{3} \sum_{z=1}^{3} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}] , C_m \} + P^{m}_{acc}$$

where the subscript "z" indicates the zone to which the call is being made with z = 4 indicating intra–LATA toll calls. The "t" subscript indicates the time–of–day rate period. The three rate periods correspond to weekday day rates (1), weekday evening rates (2) and weekday night and weekend rates (3). Other variables are

B is the household monthly telephone budget,

$p_{o,tz}$ is the cost of the first minute of a call to zone z during TOD period t,

$p_{1,tz}$ is the cost of an additional minute of a call to z during t,

y_{tz} is the number of calls made by the customer to zone z during one month of calling during t,

x_{tz} is the number of minutes of calls made during one month of calling to z during t,

C_m is the monthly credit that is deducted each month for calls in zones 1, 2 and 3, and

P^{m}_{acc} is the cost of access for measured service.

For flat rate customers the budget constraint is

$$B = \sum_{t=1}^{3} \sum_{z=2}^{4} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}] + P^{f}_{acc}$$

These can be combined into a more general formulation for non–Lifeline service:

$$B = \sum_{t=1}^{3} \delta_m[(p_{o,t1} - p_{1,t1}) \cdot y_{t1} + p_{1,t1} \cdot x_{t1}]$$

$$\sum_{t=1}^{3} \sum_{z=2}^{4} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}]$$

$$-\delta_m \cdot \mathrm{Min}\{ \sum_{t=1}^{3} \sum_{z=1}^{3} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}] , C_m \}$$

$$+ \delta_m \cdot P^{m}_{acc} + \delta_f \cdot P^{f}_{acc}$$

where δ_m and δ_f are dummy variables equal to 1 if the customer is on measured service or flat rate service respectively and 0 otherwise.

The cost of lifeline service is

$$B = \begin{bmatrix} \$0. \cdot \Sigma_t y_{t1} \\ \$.1 \cdot \Sigma_t y_{t1} \\ \$.15 \cdot \Sigma_t y_{t1} \end{bmatrix} \forall \begin{bmatrix} y_{t1} \leq 60 \\ 60 < y_{t1} \leq 70 \\ 70 < y_{t1} \end{bmatrix} + \sum_{t=1}^{3} \sum_{z=2}^{4} [(p_{0,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}]$$

Under the circle community plan all calls to zone 4 are discounted by 30 percent for a set fee, P^{cir}. This is incorporated into the budget constraint as

$$B = (1-\delta_{cir}) \cdot \sum_{t=1}^{3} \sum_{z=2}^{4} [(p_{o,z} - p_{1,z}) \cdot y_z + p_{1,z} \cdot x_z]$$

$$+ \delta_{cir} \sum_{t=1}^{3} \sum_{z=2}^{3} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,tz} \cdot x_{tz}]$$

$$+ \delta_{cir} \cdot \gamma \cdot \sum_{t=1}^{3} \cdot [(p_{o,t4} - p_{1,t4}) \cdot y_{t4} + p_{1,t4} \cdot x_{t4}]$$

$$-\delta_m \cdot \mathrm{Min}\{ \sum_{t=1}^{3} \sum_{z=1}^{3} [(p_{o,tz} - p_{1,tz}) \cdot y_{tz} + p_{1,z} \cdot x_{tz}] , \ C_m \}$$

$$+ \delta_m \cdot P^{m}_{acc} + \delta_f \cdot P^{f}_{acc} + \delta_i \cdot P^{cir}$$

where γ is the discount on calls made into zone 2 and 3 when subscribing the Circle optional Calling Plan.

MODELING THE EFFECTS OF HOUSEHOLD CHARACTERISTICS ON TELEPHONE USAGE AND CLASS OF SERVICE CHOICE

Alan J. Cox
and
Paul A. Ruud

Schools of Business Administration
and
Dept. of Economics
University of California at Berkeley

1. Introduction

Telephone companies are facing increased regulatory and competitive pressure to adjust their residential rates to cover more of the costs associated with service to that sector. The offering of a range of two-part tariffs may permit telephone companies to adjust revenues while maintaining a high degree of penetration of service. The impact of any adjustments will depend upon the distributions of household characteristics and tastes. Despite the need for such information, little work has so far been done at the household level.[1]

The lack of analysis of this problem is no doubt due to the unavailability of data. We have available a data set that includes the calling behavior, over several months, of over ten thousand telephone subscribers. The number of calls made and total duration of calls were recorded separately for each time-of-day period and for each calling zone within the subscriber's LATA. Recurring monthly charges were also recorded. The households in this sample had up to three classes of service to choose from as well as various optional calling plans.

---

[1] Train *et al.* (1987) does examine data at the household level and does include household characteristics in its specifications.

In addition to their telephoning behavior being recorded all households in this data base were contacted by telephone and asked to provide information on number of persons in the household, number of wage earners, ages of each individual and family income. Furthermore each household was asked to provide information on use of the telephone for business (including computer) and household purposes. To this data-base were added other variables that previous literature had indicated may be important to determining telephone usage, such as number of other telephone lines in the local calling area of each subscriber.

Aside from the problems of data availability, the analysis of telephone demand is made difficult by the non-convexities and non-linearities in the budget set that are introduced by the nature of the rate structure. The approach that we take to this problem is to employ the classical model of utility maximization following Burtless and Hauseman (1978).

Another important aspect of choosing amongst non-linear tariffs is that some of these tariffs have greater appeal to those who are uncertain about their own calling behavior or may be uncertain about the rate structure. Uncertainty is inherent in telephone usage. A residential phone is often used by many household members whose use is loosely supervised. The costs of monitoring even one's own phone usage imposes "demand-side transactions costs" which in turn inhibits telephone usage. These and other reasons may explain why many telephone subscribers do not choose the cost-minimizing class of service. Such a manifestation of risk aversion should be correlated with a higher degree of variability in calling behavior amongst those who display a willingness to pay more for unmeasured service. We provide some evidence of this from our data-base in this paper.[2]

In this paper we present a model of subscriber choice of class of service. The model takes into account the budget non-linearities mentioned above and incorporates uncertainty, so that we do not restrict telephone subscribers to choose the least cost service for their observed calling pattern. It also allows a relatively rich parameterization of the effects of household characteristics to explain differences in calling behavior. We also allow for the distribution of tastes.

There are other features of the demand for telephone usage that we incorporate into our model. Telephone use is described as an input to the production of "specific satisfaction" (Gorman,1976) which is the argument of the utility function that we employ. The production function that we use takes into account the observed characteristics of calling behavior, though at the expense of a good deal of complexity. This complexity makes conventional estimation unattractive and we hope to utilize simulated method of moments to overcome this problem.

It is the objective of this research effort to estimate the demand for telephone services at the household level. Once demand functions have been estimated for telephone service the information can be used to examine the effect of a variety of rate structures in the residential sector. Specific

---

[2] Wenders (1987) suggests other explanations of the preference for unmeasured service and provides a fuller discussion of those mentioned above.

issues that can be addressed include the effects on revenue of altering the time-of-day structure of rates and the question of which segments of the residential sector will be made worse off or better off by changes in the rate structure.

The following section gives some historical background to the public policy issues that we wish to ultimately address. Section 3 provides information on the rate structure facing those customers who participated in the survey. In that section we also show how the non-convexities and non-linearities arise out of the rate structure. Section 4 describes the parametric utility function that we employ in this model, its solution and its extension to take into account uncertainty. Section 4.6 gives the econometric specification of the model. In section 5 we describe the estimation method of simulated method which we hope to employ in order to estimate this model. Section 6 describes some of the evidence that we have of rational, risk-averse behavior by participants in the study.

2. Historical Background

The divestiture by AT&T of its local operating companies eliminated a large part of the funds previously used for direct cross-subsidies to residential subscriber access. In response to political pressure to keep subscriber access rates low the FCC and the state regulatory commissions required the interexchange carriers (IECs) to pay the Bell Operating companies (BOCs) a "Common Carrier Line Charge" the amount of which varied with volume of calls. This was merely, in effect, an extension of the old separations process. Some subsidy continued to be and still is derived from intra-LATA long distance rates. Higher rates for business customers were also used as a source of additional funding to subsidize local residential access.

The effect of imposing access fees upon IECs and of trying to cross subsidize residential subscriber access from commercial service is well known and documented. (See for instance, Jackson and Rohlfs, 1984). Large customers merely provided their service enhancements with customer premise equipment and bypassed the local network when making inter-LATA long-distance calls. The lost revenue to all local operating companies in 1984 from all sources of bypass was estimated to be $230 million, with the 1987 figure forecasted to be $770 million (Fischetti, 1985, p.80). Other developments, such as permitting shared tenant services, have also made unsustainable some of the cross-subsidy of services in the telephone industry.

However, the maintenance of the intra-LATA service monopoly called for in the MFJ allows BOCs to impose MTS rates upon all customers that are greater than marginal cost. This monopoly seems to be secure for the near future. The most recent rulings of Harold Greene, the presiding judge in the AT&T anti-trust case, indicate that he has no intention of permitting any major inroads to the local monopoly of the BOCs in the near future. (Telecommunications Reports, 1987).

Maintenance of the *status quo* includes continued subsidy of local access by charging for toll calls and business services at a rate greater than cost. In California, for instance, this is both implicitly and explicitly done. One explicit cross-subsidy is a PUC mandated tax on MTS calling that is earmarked to underwrite the cost of providing lifeline service. This cross-subsidizing strategy may be extended into the provision of enhanced services through "social contracts" which some PUCs are entering into with the Bell Operating Companies. Under these social contracts the BOCs undertake to maintain the current residential rate structure for a set number of years. During that period they are allowed to price business and enhanced services in any manner that they see fit. They are also allowed to keep any savings that might result from improved management or the incorporation of new technologies into the system. One element of the strategy that the BOCs and the PUCs have in setting up social contracts is to generate enough revenue from products in which BOCs have a comparative advantage in order to continue to subsidize residential subscriber access.[3]

In the future, it seems likely that the same pressures that led to divestiture will also lead to competition being introduced to the intraLATA market[4] and to the provision of enhanced services. In this event, a price structure that would allow for some cross-subsidy of local access from intraLATA toll calling or from enhanced services would become unsustainable.

Non-linear tariffs may exist that would meet the objective of maintaining a high degree of access and minimizing the degree to which residential service needs to be subsidized. Providing a range of two-part tariffs will also obviate the need to identify consumers of particular types and thus explicitly discriminate amongst them. However, a requirement for designing such a multi-part rate structures is some understanding of the distribution of tastes for telephone use. With such information the utility will know how subscribers will distribute themselves amongst the services offered and what their calling behavior will be.[5]

---

[3] For a description of such a proposal see testimony of Lee Camp to the California PUC, application 85-01-034.

[4] See the testimony of A. E. Kahn before the California PUC, application numbers 82-11-07, 83-01-21, and 83-04-02, and Huber, 1987.

3. The Telephone Rate Structure

Three basic classes of service (COS's) were available to the telephone customers that were surveyed in this study; local flat rate service, local measured service (LMS) and lifeline service. These three types of service differed almost entirely in the way local calls were charged. Under flat rate service, which has a relatively high access fee, all local calls are free. Under local measured service there was a charge for all local calls. However, the subscriber to LMS received a three dollar credit against all calls. Lifeline service was available to those who certified that they met certain means tests. For a still lower monthly access fee they received 60 free calls a month. If they exceeded this limit they were charged increasingly higher prices for telephone use. During the remainder of this paper we will primarily discuss the choice between measured service and flat rate service. Rates for all COS's are shown Table 1.

At the time that the data were collected for this study some customers could make only two types of calls for which we have data; local or toll (intra-LATA long distance). If a call was not within the local calling area it would be charged at the long distance rate. However, in two large metropolitan areas, all telephone subscribers could make calls into nearby exchanges that surrounded their local calling area at a price lower than that indicated for long-distance calls of the same distance. For each household there could be as many as two such extended calling areas, resulting in three concentric areas for local (or zone 1) calling, zone 2 calling or zone 3 calling. Calling into zone 2 and zone 3 area is referred to as ZUM calling. Not all customers who had a zone 2 calling area had a zone 3 calling area and not all customers who had a zone 3 calling area had a zone 2 calling area. The credit that subscribers on measured service receive is applied to both local calls and to ZUM calling, further complicating the explicit modeling of class of service choice.

---

[5] Brown and Sibley (1986) provides a thorough overview of two-part tariffs.

The credit has important consequences, as we will see below, since it is possible that a person who makes few local calls but $3.00 worth of ZUM calls will, in fact, save over $6.00/month on measured service relative to being on flat rate service. Complicating the issue further is the variety of access fees that are charged by the telephone company. For instance, many exchanges in small rural communities have local calling privileges beyond the distance that would normally constitute a local calling area. In return for this extended area service (EAS), which every telephone subscriber in the exchange has to take, subscribers are charged a higher price. These premiums over the usual rate are between 5 and 150 percent of the base rate, depending on the number of telephone customers in the local calling area that are being added. There are 634 customers who completed the survey who live in areas in which a premium is charged.

In addition to compulsory Extended Area Service several optional extended area services were offered, usually referred to as Optional Calling Plans. For instance, under one optional calling plan, all calls outside the Zone 2 and 3 calling areas are discounted by 30 percent for a monthly fee of $4.75. 826 customers in this survey participated in Optional Calling Plans of some sort.

Table 1

Telephone Rate Structure Facing Survey Participants

1. SUBSCRIBER ACCESS RATES

Flat Rate Access	$8.25/month
Measured Service Access	$4.45/month
Lifeline Flat Rate Service	$4.45/month
Lifeline Measured Service	$2.30/month

2. USAGE RATES

2.1 LOCAL CALLS
Flat Rate Service

Regular and Lifeline	All local calls free.

Measured Service Rate

Daytime	1st Min.:	4 cents
	Each Add. Min.:	1 cent

Lifeline

1st 60 calls:	Free
61-70 calls:	10 cents/call
Calls over 70:	15 cents/call

No Time Of Day Discounts

2.2 NON-LOCAL CALLS
Calls to Zone 2

Daytime	1st Min.:	8 cents
	Each Add. Min.:	3 cents

Calls to Zone 3

Daytime	1st Min.:	10 cents
	Each Add. Min.	5 cents

2.3 TIME-OF-DAY DISCOUNTS
All Zones, Flat and LMS

Evening	30 percent discount
Night	60 percent discount

3. CREDITS
$3.00/month measured service credit is applied to ZUM and local charges.

The rate structure for telephone services violates the usual assumption that prices are constant for all quantities. When all prices are constant and strictly positive, the budget line for a household is a straight line. Any conventional utility function maximized under a budget constraint will yield unique positive solutions to the demand for all goods. Figure 1 displays such a budget constraint with the amount of a constantly priced good q measured along the horizontal axis and the amount of all other goods c measured along the vertical axis.

The situation is much more complex under the the telephone rate structure described above. Consider, first, a situation in which there is no time-of-day pricing, only local calling, (i.e. no extended area calling and no long-distance calling of any sort,) and no charges for duration. Figure 2 shows the budget set for a telephone subscribing household under flat rate service. The horizontal axis now measures the number of local telephone calls and the vertical axis represents the value of all non-telephone consumption, c . y represents the household's monthly earnings. If p^{f}_{acc} is the cost of access under flat rate service then the continuous horizontal line at $y - p^{f}_{acc}$ represents the budget line for this household.

Under LMS the subscriber pays a lower charge for telephone access, pays a usage charge for each call and is allowed a credit of $3.00/month against calls. All these features are reflected in the solid kinked line in Figure 3. The intercept with the vertical axis is higher than for flat rate service because of the lower access fee. The horizontal segment indicates the consequence of the credit, which has the effect of paying

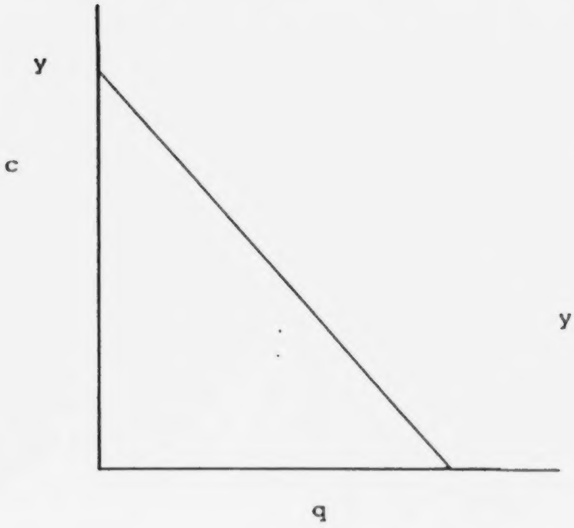


Fig. 1
Constant positive-price
Budget Line

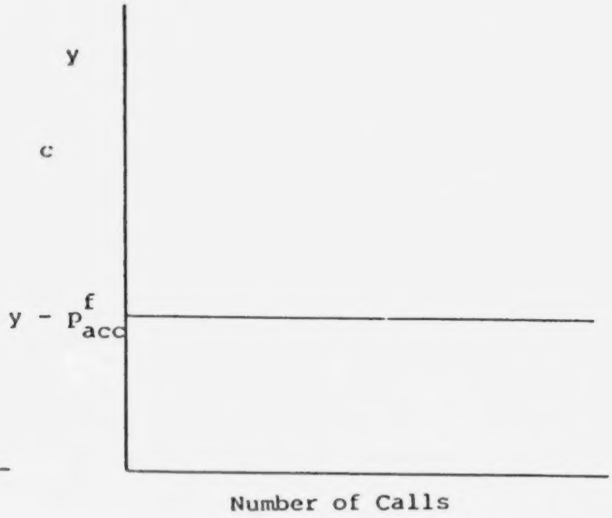

y
c
$y - p^{f}_{acc}$
Number of Calls

Fig. 2
Flat-rate Service Budget Line

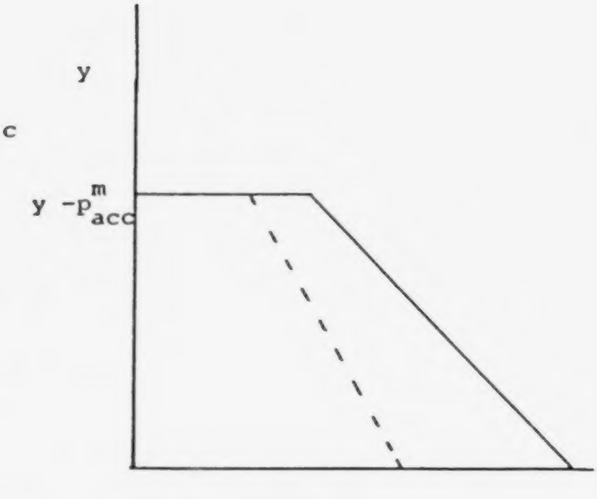


Fig. 3
LMS Budget Line

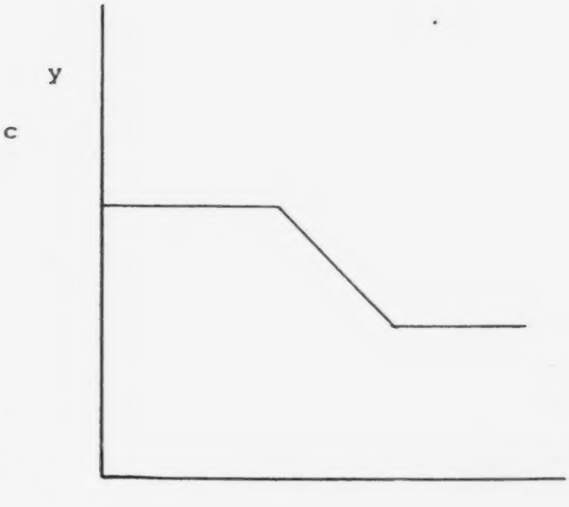


Fig. 4
Combined Flat Rate and LM Service
Budget Line

back the telephone subscriber for each call that he makes. After the credit is entirely used up the LMS subscriber pays a constant cost for all additional calls. That point is indicated by the kink in Figure 3 with the downward portion of budget line indicating the price for each call made.

The situation is made more complicated if we allow for duration charges on local measured service. In that case, the credit is exhausted with fewer calls as the average duration of calls increases. In Figure 3 the effect of increasing the average length of call is for the kink to appear along the horizontal segment of the LMS budget line but at a point to the left of the kink that appears in the solid line segment. This is depicted by the dashed line in Figure 3. The slope of the downward sloping portion of the curve would also be steeper since the cost of each call would go up with average duration.

Finally, Figure 4 depicts the potential telephone subscriber's budget set for telephone service when we consider only the dominating portions of the budget set and holding average duration of calls fixed. Now, in addition to the convex kink found in Figure 3, we also have a non-convex kink where the budget sets for LMS and flat rate calling meet at N^{**}. This makes it possible to have more than one tangency of an indifference curve with the budget line.

We have not described the situation if lifeline calling were also a possibility. Here we only note that, depending on the average length of call (and, therefore, on the position of the LMS budget set) the combination of all three options may result on two convex and two non-convex budget kinks.

The situation is further complicated when ZUM calling is allowed. The $3.00/month measured service credit is also applied against a subscriber's ZUM calling. Taking account of calls and durations for up to three zones and three time-of-day periods there is a possibility of up to 18 commodities over which the credit may have to be allocated. The behavior of most consumers are not affected by the credit since they are consuming more or less than the amount indicated by the kink. However, in modeling the behavior of a consumer who is on the kink, it is important to know how an optimizing subscriber to telephone service allocates the credit. In some cases the credit may be large enough to induce some calling where there might otherwise have been none.

4.1 A Parametric Utility Function for Telephone Usage

We will base our statistical analysis of telephone usage on a model of consumer choice derived from a parametric utility function. In contrast to Deaton (1978) and others, we will work from the *direct* utility function, rather than the *indirect* utility function. In the present case, the direct utility function is easier to motivate. The disadvantage of this approach is that the demand functions become analytically intractable and in our application we will overcome this algebraic difficulty with numerical solution methods. Indeed, the solution to the consumer's optimization problem is so awkward mathematically that the econometric specification is complex and usual methods like maximum likelihood are computationally infeasible. The numerical solutions that we will use for solving the consumer choice problem are appropriate for the simulated moments method that we will use for estimation.

Our approach to motivating a parametric utility function for telephone services is to develop equations that describe the production of "specific satisfaction" from the uses of the telephone. This specific satisfaction enters as a consumer good in the utility function of the consumer. Let t be a vector of levels of telephone services available to the consumer and let c be the value of all other goods. We will assume that the tastes of the consumer can be represented by a utility function $U(t,c)$ that is separable in t and c. Besides the reasonable character of this separability, this assumption also reflects the nature of the data that we will analyze, which does not provide a disaggregation of c. In addition, we will assume that the utility function is restricted so that

$$U(t,c) = F[\tau(t),c] , \tag{4.1}$$

where $\tau(t)$ is a production function yielding specific satisfaction. This approach has been followed by others, and in studies of telephone usage the specific satisfaction has been called *information* or *entertainment*.

In general, the vector t of telephone usage consists of the time, duration, and destination of each telephone call made over the budget period. In our case, this vector is reduced to the number of completed calls to each calling zone during specific time-of-day periods and the average duration of those calls. The zones, which we will index by z, are local zones 1, 2, and 3, and intra-LATA toll. The time-of-day periods, which are indexed by j, are day, evening, and night. Let n_{jz} be the number of completed calls to zone z during time period j and let d_{jz}

be the total duration of those same calls. The vector t is $[(n_{jz}, d_{jz})']$.

4.2 Properties of the Production Function

A production function τ is expected to satisfy several properties. We will specify a parametric functional form for τ which satisfies these properties. Here we describe and motivate the properties themselves.

Property 1: Holding the total calling time fixed, the specific satisfaction should decline as the average duration of telephone calls goes to zero. It is plain that extremely short calls do not allow the consumer to send or receive information, and that this cannot be overcome by placing more calls. On the other hand, as the duration of a call increases in the neighborhood of zero, the specific satisfaction will surely increase.

Property 2: Holding the total calling time fixed, the specific satisfaction should decline as the number of calls approaches zero (and the average duration approaches infinity). This property reflects the common sense that consumers generally prefer several calls at a choice of times and to a choice of destinations over a single call to one destination lasting the same total duration.

Property 3: Specific satisfaction should decrease after the total duration reaches a certain point. This property requires more than the usual law of diminishing returns. It captures a technological limit on the length of calls due to the difficulties of making connections and continuing conversations. The parties who are called may not be able to make or maintain the connection as long as the caller desires. The callers themselves may suffer disutility from long durations caused by fatigue.

For a parametric functional form, we propose

$$\tau(t) = \sum \rho_{jz} \, d_{jz}^{\gamma_d} \, (\mu_d^{jz} - d_{jz}) \, n_{jz}^{\gamma_n} \, (\mu_n^{jz} - n_{jz}) \,, \tag{4.2}$$

where γ_d, γ_n, ρ_{jz}, μ_d^{jz}, and μ_n^{jz} $(j=1,\ldots,J;\ z=1,\ldots,Z)$ are positive, constant parameters. Properties 2 and 3 are met by this production function, where the μ^d parameters are determinants of the limits on total calling time. Property 1 is also satisfied because specific satisfaction begins to decline when number of calls reach $n_{jz} = \mu_n^{jz}/(1+\gamma_n)$, or average duration of calls fall below $(1+\gamma_n)d_{jz}/\mu_n^{jz}$. The ρ's are weighting factors that enter into the aggregation of specific satisfaction over zone and time periods. With richer data, one might specify a richer functional form for this normalizing factor as a function of time and destination of call. Each ρ can be interpreted as the relative value of calls in each time-ZUM. We expect day-time calls to have the highest values but we cannot predict relative zone values.

Given the production function for phone services, the parametric utility function is specified to be

$$U(t,c) = \delta \ln(c) + \tau(t) \,. \tag{4.3}$$

Taken together, equations (4.2) and (4.3) generate preferences with indifference curves (for a given set of parameters) depicted in Figure 5. Note how Property 3 makes the indifference map different from textbook versions: there is a point of satiation in the duration of telephone calls labelled μ^* .

4.3 Solving the Consumer's Optimization Problem

According to the usual assumptions, the consumer chooses (t,c) to maximize the utility function (4.3) subject to a budget constraint

$$y = c + \sum p_d^{jz} d_{jz} + p_n^{jz} n_{jz} + p_{acc} , \tag{4.4}$$

where y is disposable income, p_n^{jz} is the net marginal cost of placing a call to zone z at time j , p_d^{jz} is the net marginal cost of each minute of the call, and p_{acc} is the cost of access to the telephone system. The p_d^{jz} can include both telephone charges and the opportunity cost of time so that this budget constraint can be viewed as Becker's (1965) "full income" budget constraint. However, we will treat leisure time and working hours as given, so that disposable income is constant. A positive opportunity cost of time will alter the apparent budget constraint from that depicted earlier in Figure 4 by increasing all slopes as in Figure 6.

Although one cannot derive an attractive algebraic form for the demand functions of c and t as functions of prices and income, the solution to the constrained optimization does have useful features. The solution is

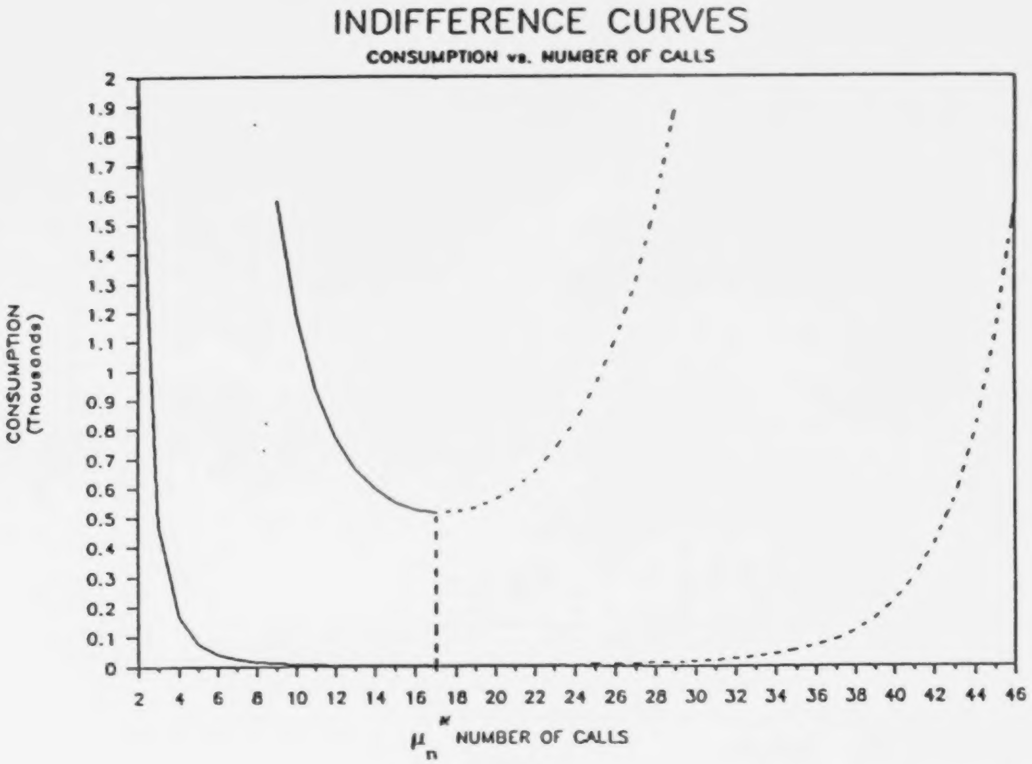


Fig. 5

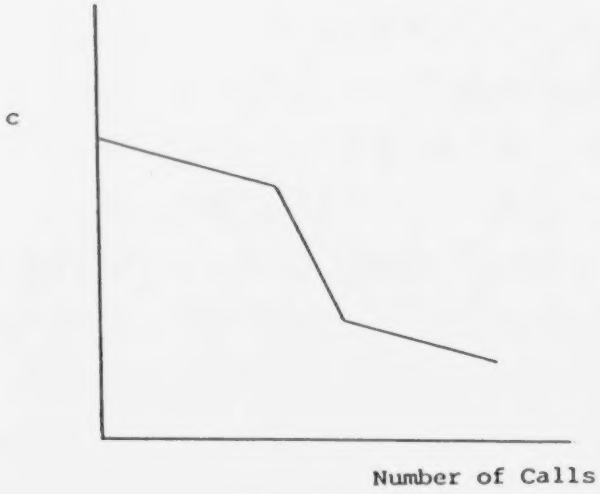

C
Number of Calls

Fig. 6
Effect of Incorporating Opportunity Cost
of Time into Full Income

particularly simple when a marginal price is zero, as in flat rate service or lifeline services:

$$p_d = 0 \Rightarrow d = \mu_d \gamma_d/(1 + \gamma_d) = \mu_d^* < \mu_d ,$$
$$p_n = 0 \Rightarrow n = \mu_n \gamma_n/(1 + \gamma_n) = \mu_n^* < \mu_n ,$$

where we drop superscripts for clarity. When both prices are positive,

$$\text{(4.5)} \qquad \frac{p_d\, d\, (\mu_d - d)}{(\gamma_d + 1)(\mu_d^* - d)} = \frac{p_n\, n\, (\mu_n - n)}{(\gamma_n + 1)(\mu_n^* - n)}$$

which permits one to find a simple and unique solution for the number of calls n as a function of the total duration d, or vice versa, with the restrictions that $n \in [0,\mu_n^*]$, $d_{jz} \in [0,\mu_d^*]$.

Additive separability also permits the solution for c as a function of any d or n when the corresponding price is positive. For example,

$$\text{(4.6)} \qquad c = \frac{\delta}{\rho} p_n\, n(d)^{1-\gamma_n} \{d^{\gamma_d} [\mu_d - d] [\gamma_n \mu_n - (\gamma_n + 1) n(d)]\}^{-1} ,$$

where $n(d)$ is the solution to (4.5) or μ_n^*, depending on p_n . The restrictions of positive marginal utilities and positive consumption levels imply a unique solution for this implicit function. Taken together, (4.3), (4.5) and (4.6) provide a single optimum for the constrained problem. Examples of a demand function for number of calls for various combinations of the parameters are depicted in Figure 7.

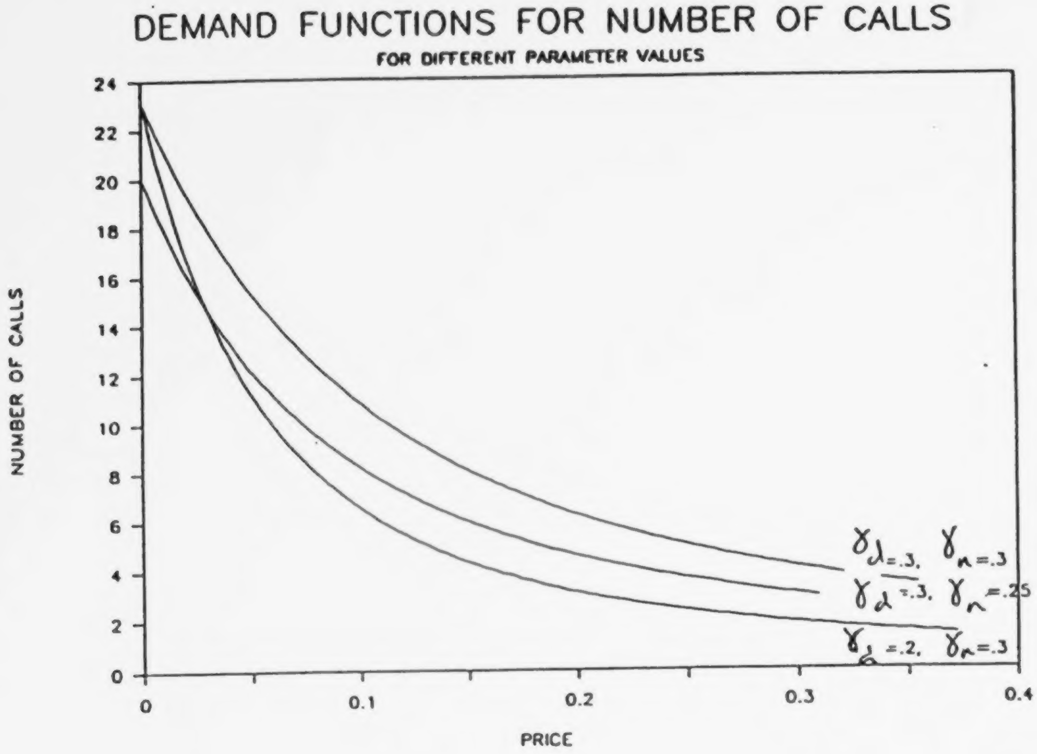


Fig. 7

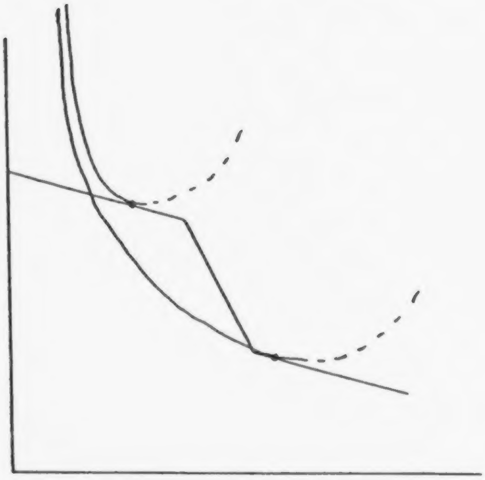


Fig. 8
Effect of Non-Convex Budget Line on
Utility Maximization

4.4 Utility Maximization over Nonlinear Budget Constraints

Our parametric specification is also motivated by a need to solve the consumer's problem in the face of nonlinear budget constraints, not just the simple linear form given in (4.4). As described previously, telephone customers face two basic choices of service, flat rate and measured, with different access fees and marginal prices, and measured service provides a credit that gives some calls a marginal price of zero. From a purely conceptual point of view, the optimal choice between services is easily found by maximizing utility for each choice and choosing the service with the highest maximized utility. An example of such a utility maximum is illustrated in Figure 8, in which measured service is preferred. From a computational point of view, such optimization requires great care given the nature of the constraints that must be satisfied and the multi-dimensional character of the goods. A typical approach is to solve several constrained optimization sub-problems.

An important special case arises due to the credit in measured service (and a rising marginal cost in life-line measured service) that puts convex kinks in the corresponding budget sets. When the utility maximum lands on these convex kinks, another constrained optimization problem arises. Even in the simplest case of one time-ZUM, this appears as the problem of finding the utility-maximizing combination of number of calls and total duration along the chord AB in Figure 9.

Fortunately, maximizing utility subject to the budget constraint (4.4) is almost identical to maximizing utility subject to consuming on a budget

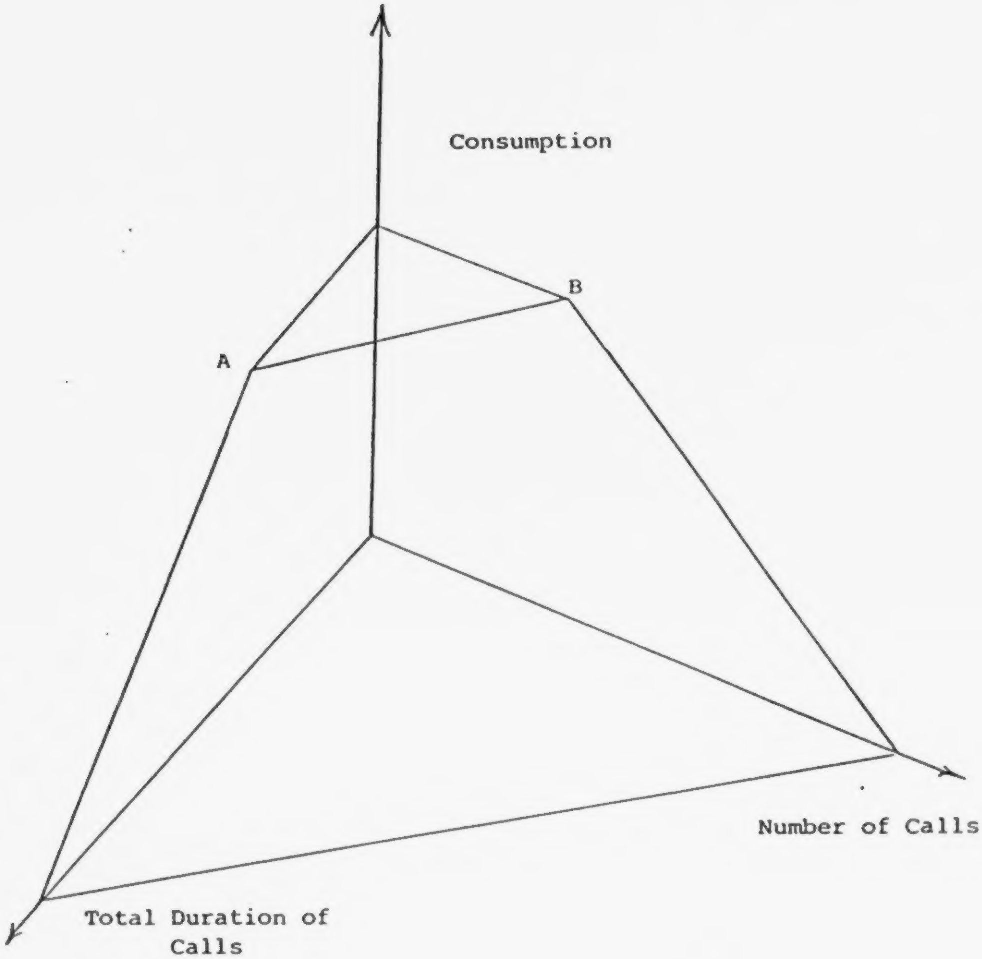

Consumption
B
A
Number of Calls
Total Duration of Calls

Fig. 9
Effect of Credit on Consumption of Calls and Duration
For A Single Time-ZUM

kink. The budget kinks have the property that c is constant. This property and the separability of the utility function effectively remove this variable from the optimization problem. Furthermore, the constrained solution continues to satisfy (4.5). As a result, an algorithm which solves the original problem in the previous section will also solve the present one after two simple modifications are made: (1) δ is set to zero, removing c from the utility function and (2) disposable income is replaced by disposable income minus the fixed level of c changing the constraint (4.4) to

$$(4.7) \qquad y - c = \sum_{j,z} p_d^{jz} d_{jz} + p_n^{jz} n_{jz} + p_{acc} .$$

4.5 Decisions Under Uncertainty

We additionally want our parametric model to allow for the presence of uncertainty in the consumer's problem. Uncertainty is inherent in telephone usage. Telephones are frequently used to communicate unanticipated information as it comes to light. A residential phone is often used by many household members whose use is loosely supervised. In addition, the costs of monitoring phone usage inhibit such activity and users do not know their current telephone expenditures. The influence of such sources of uncertainty on consumer decisions may explain the large fraction who subscribe to flat rate service even though measured service appears to be less costly.

Uncertainty is incorporated in our parametric model by ascribing distributions to some of the parameters. We are interested in capturing the effect of this uncertainty on the choice of service, and not on calling behavior given the service choice. Let the collection of μ's have a joint multivariate distribution function $F(\mu,\theta)$ depending on parameters θ, which is known by the consumer. In choosing their service option, consumers are assumed to maximize expected utility in determining their telephone service and calls:

$$\begin{aligned} (4.8)\quad E[U(c,t)] &= \int U(c,t)\, dF \\ &= \delta\, ln(c) + \sum \rho_{jz}\, d_{jz}{}^{\gamma_d}\, n_{jz}{}^{\gamma_n}\, E[(\mu_d^{jz} - d_{jz})(\mu_n^{jz} - n_{jz})] \\ &= \delta\, ln(c) + \sum \rho_{jz}\, d_{jz}{}^{\gamma_d}\, n_{jz}{}^{\gamma_n}\, (\theta_{\cdot}^{jz} - \theta_n^{jz} d_{jz} - \theta_d^{jz} n_{jz} + d_{jz} n_{jz}) . \end{aligned}$$

In this case, the θ parameters capture covariance among the μ. The expected utility function does not differ substantially from the utility function itself. The fact that $\theta_{\cdot}^{jz}$ is not constrained to equal the product $\theta_d^{jz}\theta_n^{jz}$ does not represent a significant difference in predicted behavior. We could just as well remove this constraint in the direct utility function. The solution methods outlined above extend to this case.

This modeling of uncertainty simply shows that the parametric utility function that we propose accommodates the effects of uncertainty easily. Our model does not restrict consumers to choose the least cost service for their observed calling pattern, but allows them to prefer an option that provides insurance against large telephone bills.

Given their service, whether flat rate or measured, consumers work out their demands for calls during each billing period. We suppose that the uncertainty present in the choice of service has been removed by the time consumers are making their calls. The number of calls and their duration will depend on the net marginal cost of each call and the needs that are realized in the parameters of the utility function. In this way, consumers who appear to be identical will exhibit different calling behavior according to the service that they have chosen.

4.6 Econometric Specification

Finally, let us describe our generalization of this parametric model to accommodate heterogeneity in the population of consumers. Observable differences among telephone customers will enter into the specification of the μ parameters. Although rich parameterizations are available, we seek to be parsimonious and specify

$$(4.9) \qquad \mu_n^{jz} = \alpha_{1n}^{jz} + \alpha_{2n}^{jz}\cdot x'\beta_n + \varepsilon_n^{jz} ,$$

$$(4.10) \qquad \mu_d^{jz} = \mu_n^{jz}\cdot exp(\alpha_{1d}^{jz} + \alpha_{2d}^{jz}\cdot x'\beta_d + \varepsilon_n^{jz}) , \quad \mu_n^{jz} > 0 ,$$

so that an index for numbers of calls $x'\beta_n$ and an index for duration $x'\beta_d$, that vary with observable characteristics x , determine all of the μ's. The ε's are random variables that contain the uncertainty in the μ's described in the previous section. The exponential transformation of duration conveniently restricts duration to be strictly positive, as it should be. To account for the variation in duration that is attributable to the number of calls, we specify a proportional relationship in (4.10).

In our econometric specification, we also allow for unobserved heterogeneity in the form of a distribution of tastes over the population of consumers. Thus, the ε disturbances in (4.9) and (4.10) are also given the role of capturing this variation in tastes, besides their role in modeling uncertainty. An error components structure delineates the taste component, which the consumer observes, from the uncertainty component, which is observed after the choice of service:

$$(4.11) \qquad \varepsilon_n^{jz} = \eta_n^{jz} + \xi_n^{jz} \quad \text{and} \quad \varepsilon_d^{jz} = \eta_d^{jz} + \xi_d^{jz} ,$$

where the η's are the taste component and the ξ's are the uncertainty component.

The distribution of the components of ε_d is specified conditional on $n > 0$, and hence ε_n, because duration is not defined for zero calls. Because the number of calls must be positive and is frequently zero, the marginal distribution of ε_n must be a censored one, with a nonzero probability of zero calls. We specify a mixed density with a continuous distribution for positive calls. While it is surely necessary to capture the censoring at zero, neglecting the integer character of strictly positive calls seems reasonable. From the consumer's perspective, (η_n, η_d) is observed so that the distribution of the components of uncertainty (ξ_n, ξ_d) is logically conditioned on the η's.

For these reasons, we specify the marginal distribution of η_n to be a censored normal so that

(4.12) $$\mu_n^{jz} - \xi_n^{jz} = \alpha_{1n}^{jz} + \alpha_{2n}^{jz}\cdot x'\beta_n + \eta_n^{jz} \geq 0 \, .$$

The conditional distribution of η_d for strictly positive $\mu_n - \xi_n$ is normal. The conditional distribution of ξ_n is also a censored normal so that

(4.13) $$\mu_n^{jz} = \alpha_{1n}^{jz} + \alpha_{2n}^{jz}\cdot x'\beta_n + \varepsilon_n^{jz} \geq 0 \, ,$$

so that no direct dependence on η_d is specified. Analogous to η_d, the conditional distribution of ξ_d for strictly positive μ_n is normal. Collecting terms together we have the following stochastic specification: let

(4.14) $$x'A_n = [\alpha_{1n}^{jz} + \alpha_{2n}^{jz}\cdot x'\beta_n] \, , \quad \eta_n = [\eta_n^{jz}] \, ,$$
$$\eta_d = [\eta^{jz}] \, , \quad \xi_n = [\xi_n^{jz}] \, , \quad \xi_d = [\xi_d^{jz}] \, .$$

then

(4.15) $$\eta_n \sim max[N(0,\Omega_n), -x'A_n] \, ,$$

(4.16) $$\eta_d | \eta_n > 0 \sim N(0,\Omega_d) \, ,$$

(4.17) $$\xi_n | \eta_n, \eta_d \sim max[N(0,\Sigma_n), -x'A_n - \eta_n] \, ,$$

(4.18) $$and \quad \xi_d | n>0, \eta_d \sim N(0,\Sigma_d) \, .$$

Each maximum in (4.15) and (4.17) is taken element by element. When a conditioning inequality in (4.16) and (4.18) is not satisfied, then the corresponding element of the random vector is not observed. Each covariance matrix, Ω_n , Ω_d , Σ_n , and Σ_d , is a positive semi-definite matrix. For parsimony we will consider restrictions on these covariance matrices. Because differences in variances are an important part of our story, we will probably want to specify some form of heteroskedasticity, either parametric and/or identified by the time-series part of the panel.

Heterogeneity is also reflected in differences in the net marginal prices of the duration of calls. The net marginal price is a function of both the telephone rate structure, which assigns a positive price for measured service and a price of zero for flat rate service, and the personal opportunity cost of time. Thus,

$$p_d^{jz} = r_d^{jz} + w_j ,$$

where r is the telephone charge rate and w_j is the opportunity cost of time in period j . We specify

$$w_j = x'\zeta_j$$

for the three time periods that are given in our data. A distribution for the w_j is also attractive, but we have not ascribed one. An unusually large cost of time, for example, is an explanation for no calls in a particular time-ZUM. Note that we do not constrain w_j to be positive, reflecting the possibility that phone calls may augment full income.

Two other items remain to be discussed. Firstly, we could well specify the ρ's to be random variables. This is attractive as a means of capturing variation in elasticities, freeing up the burden of explanation on the μ parameters as limits of calling behavior and as central tendencies. The ρ parameters also offer another way of capturing covariance among calls. We will typically normalize the δ parameter to unity for identification. Secondly, we recognize that additional variation in the data comes from deviations from maximizing utility as we have set this up. This topic arises naturally in the context of estimation as errors in the specification of moments.

5. Simulated Moments Estimation

The estimation of this econometric model by maximum likelihood is an unattractive prospect because many of the numerical calculations that are required are difficult to carry out on current computers. The computation of the probabilities of choosing each service, for example, involves *multivariate* integration over complicated regions that are only defined *implicitly*. We have chosen to be ambitious in our specification of the model at the cost of such convenience in estimation. Our hope is to address a potentially complicated estimation problem with the method of simulated moments proposed by McFadden (1987) and Pakes and Pollard (1987).

In the estimation method of simulated moments, one chooses parameter estimates by minimizing the distance between observed data and simulations of that data as predicted by the parametric model. In the present case, we observe the choice of service and the number and the duration of calls in twelve time-ZUM's. Denoting the choice of service by the variable s_1, which is 1 if consumer i chooses flat rate and 0 if consumer i chooses measured service, the vector of numbers of calls by consumer i in each time-ZUM by n_i, the corresponding vector of durations of calls by d_i, we denote the entire vector of observed dependent variables by

$$y_i \equiv [s_i, n_i', d_i']' .$$

We collect all of the parameters (γ's, α's, β's, Ω's, Σ's, and ζ's) in the vector θ, denote a simulation of the ξ's and η's for consumer i by the vectors $e_i(\theta)$ and $h_i(\theta)$, and denote the predicted values of y_i, as functions of e_i, h_i, θ, and x_i (characteristics of the consumer) by $\bar{y}_i(e_i, h_i, \theta, x_i)$. If the model that we have specified were the actual mechanism generating the consumer data, then

$$E[y_i - \bar{y}_i(e_i, h_i, \theta, x_i)] = 0$$

when θ is equal to its value in the population. In that case, both y_i and $\bar{y}_i$ are being generated in the same way. At arbitrary values for θ the simulated $\bar{y}_i$ will have a different expectation. The equality at the population value for θ becomes a means for estimating that value. Using functions of the x_i as instrumental variables $w(x_i)$, a set of equations forming an implicit function for all the elements of θ can be constructed with all the observations in a data set covering N consumers:

$$\sum_{i=1}^{N} w(x_i)\, [y_i - \bar{y}_i(e_i, h_i, \hat{\theta}, x_i)] = 0 .$$

This simple form for a simulated moments estimator can be improved for estimator efficiency by judicious choice of the instrumental variables.

This method of moments approach to estimation is particularly appropriate for allowing more variation in the data than our original specification of the model captures. Deviations in behavior from the predictions of the model are accommodated by estimating parameter values that fit the observations as closely as possible. In terms of data generating mechanisms, the simulated moments estimator $\hat{\theta}$ is a consistent estimator provided that the simulations are unbiased predictors of the y's, which can happen even when the specified model does not exhibit all the variation present in the data. From a geometric point of view, $\hat{\theta}$ is minimizing the distance between y and $\bar{y}$, where the outer-product of the instruments serves as the measure of distance:

$$\hat{\theta} = \operatorname{argmin} \; [y - \bar{y}(\theta)]' W W' [y - \bar{y}(\theta)] \; .$$

6. COS Choice and Uncertainty

In Section 3 we showed the manner in which the telephone rate structure resulted in both convex and non-convex kinks in the budget set, even when we consider only flat rate service and LMS. In section 4 we described the manner in which these problems could be taken into account in describing the class of service choice under both certainty and uncertainty. In this section we describe some results from an examination of our data that indicate that subscribers are choosing their class of service both on the basis of cost and risk.

Figure 10 illustrates another phenomenon of non-convex budget segments. In this case, a single indifference curve can be tangent with the budget set at two locations. The curve's location illustrates a well-known property of two part tariffs: nobody who correctly anticipates the amount of calling will intend to make the number of calls indicated by the region around the kink-point at N^*, no matter what class of service choice they make. A person with such a utility function would either choose LMS and make only N^1 calls *or* flat rate service and make N^2 calls. To the extent that persons are observed close to non-convex kinks it is necessary to model a relationship between expected number of calls and actual calls as we have done above.

The region around N^* is actually part of a wider area where intentions are clearly at odds with actual behavior, depending upon class of service chosen, and for which the uncertainty framework developed above is necessary. For instance, any household on measured service could not have expected to be making more than N^1 calls and no household on flat rate service could have expected to be making less than N^2 calls.

As reported above and by other researchers,[6] we do find a great many subscribers whose usage is not consistent with their class of service choice when we consider only cost-minimizing behavior. This is particularly true for those on flat rate service. That is, most subscribers on flat rate service would be cost-minimizing better off if they were on measured service. We have already discussed several reasons

---

[6] See, for instance, Hauseman, Tardiff and Baughcum (1986).

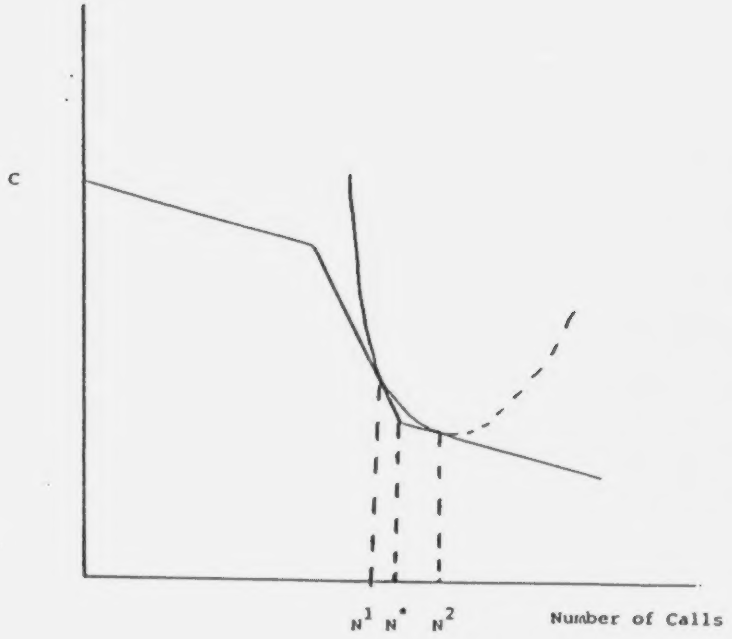

C
N^1
N^*
N^2
Number of Calls

Fig. 10

Nested Indifference Curve in a Non-Convex Budget Set why we would expect to find such a situation.

In order to undertake a preliminary examination of this question, eight months of data on household calling behavior were studied. For this data set, only households that did not alter their class of service during those eight months were studied. For each customer and for each time-ZUM rate the mean number of calls, variance of number of calls and of durations and the skewness of calling and duration were calculated. The means of number and duration of local and ZUM calls were utilized to calculate the resulting average bill under either flat-rate or LMS calling. We then calculated the saving from being on LMS. Let CF be the size of the telephone bill under flat rate service and CM be the equivalent figure under LMS. Subtracting the cost of being on measured rate service from that of being on flat-rate service gives a cost saving from being on measured service for household i as

$$CF_i - CM_i = P^f_{acc} - P^m_{acc} - \overline{LCLZMBILL}_i$$

where

$$\overline{LCLZMBILL}_i = \frac{1}{8} \sum_{n=1}^{8} (LOCREV^n_i + USECRED^n_i)$$

$LOCREV^n_i$ is the individual's cost of local calls during month *n* if all calls were being billed at the rate for LMS and before the credit was deducted,

$USECRED^n_i$ is the amount of the individual's credit that is used in offsetting local and ZUM calls,

P^f_{acc} is the cost of access under flat rate service,

P^m_{acc} is the cost of access under measured service.

Figure 11 shows the distribution of average actual or potential savings from being on measured service. The vertical axis indicates the proportion of observed subscribers in a particular rate class whose savings on measured service are at least as small as the corresponding figure on the horizontal axis. Positive values along the horizontal axis indicate a household that would have been cost-minimizing better off by being on measured service. The line that is everywhere furthest to the left is the distribution of flat rate customers. The difference between the distributions is apparently large and the difference between their means is significant at the .0001 probability level using a two-tailed t-test with different underlying population variances. The number of flat rate customers who are actually (cost-minimizing) better off under flat rate service is 43.0 per cent. On the other hand, only 10.1 per cent of measured service customers would have been better off under flat rate service.

Those on measured service are thus far more likely to have made the correct service-choice decision. This is, no doubt, partly due to the fact that most customers would be better off under measured service. On the other hand, there is also a considerable difference in the cost of a mistake when one compares the cost of making too few calls when one is on flat rate service (at the most $6.40) and too many when one is on LMS.

Table 2 indicates the average standard deviations of the number of calls per month and of the total amount time spent on the telephone for calls placed into the local calling area and into zones 2 and 3. All the customers included in this table would have had lower bills had they been

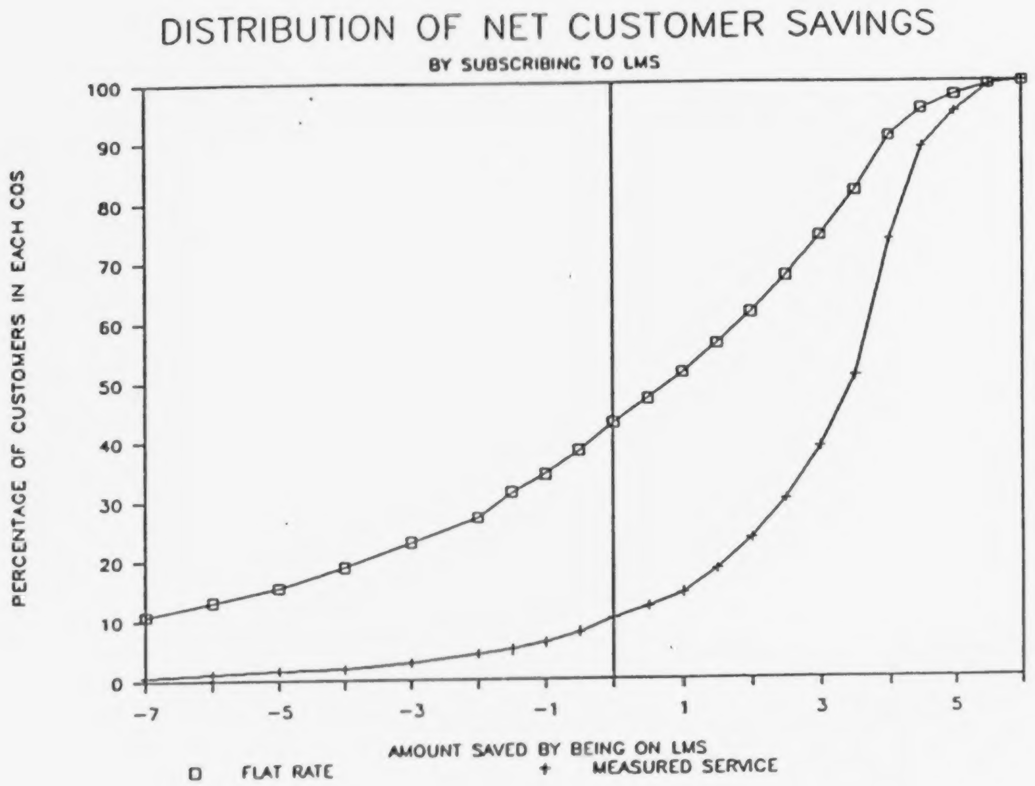

DISTRIBUTION OF NET CUSTOMER SAVINGS
BY SUBSCRIBING TO LMS
PERCENTAGE OF CUSTOMERS IN EACH COS
100
90
80
70
60
50
40
30
20
10
0
-7
-5
-3
-1
1
3
5
AMOUNT SAVED BY BEING ON LMS
□ FLAT RATE
+ MEASURED SERVICE

Fig. 11

on measured service. The first column of results are for those customers who were on flat rate service, while the second column is for those customers who made the "correct", cost-minimizing choice to be on measured service.

Table 2

Means of Standard Deviations of Monthly Calling
Of Subscribers for Whom Measured Service
Is the Least Costly Telephone Service

		Standard Deviations	
Number of Calls		Flat Rate Customers	LMS Customers
Local	Day	20.0	14.82
	Evening	13.7	10.12
	Night	15.0	11.15
Zone 2	Day	2.7	2.28
	Evening	1.5	1.55
	Night	1.8	1.68
Zone 3	Day	2.2	1.76
	Evening	1.4	1.24
	Night	1.5	1.32
Duration of Calls			
Local	Day	81.3	57.69
	Evening	78.0	60.46
	Night	70.5	53.19
Zone 2	Day	9.5	7.09
	Evening	11.9	12.58
	Night	10.6	15.71
Zone 3	Day	9.7	7.52
	Evening	10.0	9.76
	Night	9.0	7.59

Number of observations = 4,004.

The differences in the means is significant at greater than 99 per cent for all local calling figures (i.e. number and duration of calls) but are insignificant for all the other types of calls or durations. Flat rate customers tend to have greater variability in their calling patterns, thus increasing the probability that they will have unexpectedly large bills from local calling and making it more difficult to accurately estimate of the expected amount of calling.

The results of examining the eight months of data indicate, first, that potential subscribers who make a greater than average number of local and ZUM calls are more likely to subscribe to flat rate service. Considering the small cost of making an error on the side of caution, it is encouraging that this tendency is measurable at all. Secondly, those flat-rate subscribers who should, in a cost-minimizing sense, subscribe to measured service tend to have greater variability in their calling patterns than do measured service customers who have made the cost-minimizing class-of-service choice.

7. Conclusions

This paper has presented a broad discussion of some of the technical and public policy issues related to telephone pricing. We have presented a model of telephone use and class of service choice. This model takes into account many of the observed characteristics of telephone behavior in the residential sector. While complex, it can be solved through numerical techniques. The complexity of the resulting econometric specification will require the use of simulated method of moments estimation.

The work that is being proposed here is part of a general effort to increase the understanding of the demand for telephone services and the costs of providing those services. The need for this sort of information has increased significantly since the Modified Final Judgment upset the relatively quiet process of telephone regulation and the previous structure of cross-subsidies. For regulators an improved understanding of the demand for telephone services in the residential sector is necessary for effective regulation and for the achievement of public policy goals. For Bell Operating Companies a complete understanding of the nature of demand is essential to the strategic decisions of new product introduction and pricing. In addition, the work is of interest in itself, dealing with difficult issues of estimation in the face of non-linear and non-convex budget sets, risk aversion and the recursive nature of class of service choice.

References

Becker, G. S. (1965) "A Theory of the Allocation of Time", *The Economic Journal,* Vol. 75, pp. 493-517.

Brown S.J. and D. S. Sibley A. "The Theory of Public Utility Pricing" Cambridge University Press, New York, 1986.986.

Burtless, G. and J. A. Hauseman, (1978) "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment", Journal of Political Economy,Vol. 86, no. 6, pp. 1103- 1130.

Deaton, Angus, (1978) "Specification and Testing in Applied Demand Analysis," The Economic Journal, Vol. 88, pp. 524-536, Sept.

Fischetti, M. A. "By-passing is Big Business for Business," IEEE Spectrum, Nov.1985, pp. 7883.

Hauseman,J., T. Tardiff and A. Baughcum (1986) "The Demand for Optional Local Measured Service" MSU Public Utilities Papers, pp. 536-56.

Huber, Peter W., (1987) (1987) "The Geodesic Network: 1987 Report on Competition in the Telephone Industry," Report issued by the U.S. Dept. of Justice, Antitrust Division.

Jackson, C. L. and J. H. Rohlfs (1984), "Access Charging and Bypass Adoption, A Study Prepared for Bell Atlantic," Shooshan and Jackson, Inc., Washington, D. C.

Kahn, A. E. (1983) Testimony 82-11-07, 83-01-21, 83-04-02 before the California Public Utilities Commission.

McFadden, D. (1987) "A Method of Moments For Estimation of Discrete Response Models Without Numerical Integration", Unpublished.

Pakes, A. and D. Pollard (1986) "The Assymptotic Distribution of Simulation Experiments" forthcoming in *Econometrica*.

Telecommunications Reports (1987b) "Finding no Major Changes in Exchange Bottleneck or in FCC regulation, Greene says Core Restrictions Must Stand;..." and "Customer Bypass has not Ended Local Exchange Bottleneck..." Vol. 53, No. 37, pp.1-5 and pp. 5-6, Sept. 14.

Wenders, J. T. (1987) "The Economics of Telecommunications. Theory and Policy" Ballinger, Cambridge, Mass.

UNIVERSITY OF CALIFORNIA, BERKELEY

Department of Economics

Berkeley, California 94720

Working Paper 8879

REGULATION GAMES

Richard J. Gilbert and David M. Newbery

June 16, 1988

Key words: regulation, utilities, repeated game

Abstract

We examine regulation as a repeated game between a regulator and a utility facing a Markovian sequence of demands. Sunk capital would be expropriated by a regulator concerned only with the short-run interests of consumers. There exist rate of return regulatory policies supporting efficient investment paths with zero expected profits as subgame perfect Nash equilibria, but these policies must under-reward capital in some states of the world. Carefully designed nonlinear price regulation can improve on these equilibrium outcomes, although at higher consumer cost, and only if state-contingent transfers are feasible.

JEL Classification: 613, 026

Regulation Games

by

Richard J. Gilbert and David M. Newbery[1]

University of California at Berkeley

June 16, 1988

I. Introduction

The centerpiece of regulation for investor–owned public utilities is the rate of return model. The model derives its legitimacy from a series of judicical precedents, which prescribed that a regulated utility is entitled to a "fair return upon the value of that which it employs for the public convenience" (*Smyth v. Ames*, 1898). Despite a distinguished lineage, the rate of return model has been soundly criticized from the perspective of both normative and positive economics. Its incentives to choose production techniques that are too capital intensive when the rate of return exceeds the cost of capital have been documented by Averch and Johnson (1962), Baumol and Klevorick (1970) and others. Rate of return regulation has obvious allocative distortions that result from setting prices in approximate correspondence with average and not marginal costs,[2] and the "cost–plus" characteristic of rate of return regulation is generally inconsistent with efficient behavior under asymmetric information.[3] Joskow (1974) has questioned the accuracy of the rate of return regulation model as a description of actual regulatory practice. He noted that when utilities experienced rapid improvements in productivity, regulators tended to set prices, not rates of return, and that rate of return regulation emerged only in response to adverse conditions of cost and demand.

---

[1] Written while Newbery was visiting from the University of Cambridge, England. We are indebted to Eddie Deckel, Joe Farrell and Drew Fudenberg for their helpful comments.

[2] See Gilbert and Henly (forthcoming) for a discussion of the welfare costs of average cost pricing in electric power utilities.

[3] The cost–plus characteristic of rate of return regulation emerges only as an extreme case in the incentive compatible contract design analyzed by Laffont and Tirole (1986). However Besanko (1985) identifies aspects of rate of return regulation in an incentive–compatible mechanism with imperfect information about the cost of capital.

Most of the theoretical literature on public utility regulation has studied regulation in a static context or in a finite game with asymmetric information.[4] Goldberg (1976) argued that this approach overlooks a critical feature of regulation, which is its role in defining the boundaries of conduct in a continuing relationship. This view, which is consistent with Williamson's (1976) persepectives on regulation in the cable TV industry and his criticism of the auction approach suggested by Demsetz (1968), informally characterizes regulation as a repeated game. It is well known that cooperative strategies that cannot be supported as equilibria when parties are concerned only with their immediate payoffs can be supported as equilibria in an infinitely repeated game.

This paper models a continuing regulatory relationship. The emphasis is on costs which are specific to the regulated firm, and therefore sunk after investment occurs. Sunk investments can be exploited by regulators whose concerns are dominated by consumer welfare in the short–run, but if such behavior were anticipated, the firm would be reluctant to invest in further capital, to the possible detriment of the regulator's constituency. In some circumstances this state of mutual fear can be sufficient to support an equilibrium in which both parties behave in a responsible manner.

In the years between 1940 and 1970, the electric power industry enjoyed declining costs of production and steadily increasing demand. In this stable environment, it is not difficult to envision how an implicit regulatory contract might have evolved that encouraged efficient behavior despite the clearly adverse incentives of rate of return regulation.[5] In the

---

[4] Averch and Johnson (1962), Baumol and Klevorick (1970) and Bailey (1973) are examples of the former. The latter include, among other contributions in a rapidly growing literature, Baron and Meyerson (1982), Sappington (1983), Guesnerie and Laffont (1984), Riordan (1984), Laffont and Tirole (1986), Sappington and Sibley (1986), Baron and Besanko (1987), and Spulber (1988).

[5] One characteristic of this efficient implicit contract may have been the suspension of rate of return regulation in favor of price regulation, as Joskow (1973) suggests. However, rate reviews in keeping with the principle of a fair rate of return on invested capital were not

1970s a combination of higher than expected costs and lower than expected demand upset the pattern of the post–war period, and many have argued that the traditional regulatory compact has fallen into disrepair. In order to capture how an implicit contract might function in a changing environent, we consider a market in which demand follows a Markovian process, with high demand more likely if demand was high in the recent past. The efficient level of utility investment is a function of the likelihood that demand is high or low. We find that in such an environment an institution such as rate of return regulation can be first best for the regulator. However, the changing pattern of demand can cause a system of rate of return regulation to be unsustainable as an implicit contract.

The equilibria that can be sustained in a continuing regulatory relationship depend on the consequences that would follow if parties acted in their short–term interests. These in turn depend on institutional constraints that define the actions which are permissible under the law. Our main objective in this paper is to study how insitutional constraints might affect the equilibria that are sustainable in a continuing regulatory relationship. We suppose that there is a third party, a "legislator", who specifies the permissible actions of the regulator before the game begins. The regulator may choose any compensation schedules that satisfy the constraints imposed by the legislator, and we suppose that the regulator will choose the actions that maximize its expected utility. To simplify matters, we assume that the regulated firm behaves as a Nash follower, with rational expectations as to the actions of the regulator.

Section II describes the structure of technology and demand in the model and characterizes efficient investment paths. Demand is stochastic and serially correlated. If current demand is high (a state we label "good" because it would ordinarily imply high investment), it is more likely to be followed by high demand in the next period than if current demand were low (a "bad" state). The utility must invest before demand is realized, knowing

---

completely abandoned in this period, and, as we shall argue, rate of return regulation need not be inconsistent with efficient investment in a continuing relationship.

only demand in the previous period and the past behavior of the regulator. The regulator choses a compensation after the utility invests and demand is realized. If the game ended at that point, it would have a unique Nash equilibrium in which the regulator pays only marginal cost (expropriates all of the utility's sunk capital) and the utility does not invest.

Rate of return regulation loosely specifies that a utility is entitled to a fair return on its invested capital, provided that this capital was invested prudently and is employed in the public interest. In our model we interpret this institutional constraint as requiring a return at least as large as the cost of capital on all investment which is not idle. Following legal terminology that was applied to the regulatory treatment of excess capacity and cancelled plants, we refer to this rule of behavior on the part of the regulator as "used and useful" rate of return regulation (UUROR).[6]

We show in Section III that a policy of "used and useful" rate of return regulation can overcome the Averch–Johnson effect inherent in a regulatory policy that compensates the firm for its entire invested capital. There exist UUROR policies that encourage a price–taking utility to choose an efficient investment path with zero expected profit. Thus the recent behavior of regulators in disallowing expenditures on unnecessary plants or "imprudent" investments is not inconsistent with efficient regulatory behavior. The absence of such a policy would encourage either too much or too little investment.

Whether or not an efficient UUROR policy can be sustained depends on the "threats" that are permissible. Section IV considers two possible threat strategies. In the first game the regulator can and would revert to the Nash equilibrium of the one–shot game if the utility invested less than the capital required to meet baseload demand. We show that such a strategy is sufficient to sustain efficient behavior (with zero profits), with reasonable values for the cost of supply shortages. In addition, we show that a UUROR policy is efficient for the largest set of parameter values when the utility is allowed to earn strictly positive profits when demand is high, and is forced to earn strictly negative profits (on excess capacity) when demand is low. At the same time, the expected profits of the firm can be held to zero. This pattern of profitability is not inconsistent with the observed history of regulation, in which utility market values substantially exceeded their book values in the "golden years" from WWII until the mid–1960s, and then declined sharply until the mid–1980s.

---

[6] The $2.25 billion default of the Washington Public Power Supply System (WPPSS) could be construed as an example of the "used and useful" doctrine, although the legal wrinkle in WPPSS was that the member utilities were not held responsible to pay for electricity that was not delivered. Actual application of the used an useful doctrine has varied considerably in different regulatory jurisdictions. See Howe (1983) for a survey of regulatory treatment of cancelled plants.

A difficulty with the game described in Section IV is that the one–shot Nash strategy implies appropriation of sunk capital expenditures, which could be considered illegal under UUROR (although in equilibrium the utility would not invest and there would be no capital to appropriate). Thus in Section V we consider a variation of this game in which the regulator is never allowed to pay less than the cost of capital that is actually used. For reasonable parameter values, the equilibrium in this game has the utility investing only to produce base load demand. When UUROR is interpreted broadly to apply to all threatened behavior that may be feasible for the regulator, it becomes less likely that efficient investment behavior can be supported as an equilibrium.

In Section V we relax the judicial constraints imposed by the Legislator and suppose that the regulator can choose any compensation schedule in any of the four states, corresponding to demand in the current and preceding periods, but again subject to no expropriation of capital that is fully employed. We describe this alternative as price regulation. Price regulation includes UUROR as a special case, so it can be no worse and could be better. However, if the regulator is restricted to set only *linear* prices in each state, the regulator would have to transfer positive expected profits to the firm in any situation where the efficient investment path could be sustained as a SPNE under price regulation, but not under UUROR. Thus, with linear price regulation, any advantage over UUROR comes at the cost of transferring positive expected profits to the firm. In addition, if the regulator sets linear prices that depend only on the *current* state of demand, the firm must earn strictly positive profits in any efficient equilibrium. Rate of return regulation is equivalent to (nonlinear) price regulation in which prices depend on demand in both the current and the preceding period. When price regulation is restricted to policies that depend only on current realized demand, price regulation can be more expensive for a regulator than rate of return regulation because the utility must be compensated for the risk that it will have idle capital that does not earn revenues.

II The Model

In each period t the utility chooses a non–negative level of investment, k_t, to bring total capital in the next period up to the desired level, K_{t+1}. The period is sufficiently long that capital completely depreciates if there is no new investment, so that $K_{t+1} = k_t$. At the time of the investment decision, demand in period t+1, D_{t+1}, is not known, but may take either of two levels (both independent of price over the relevant range), Q or $Q(1-\sigma)$. (The inelasticity of demand simplifies the analysis, but suppresses questions of allocative efficiency). Demand follows a Markovian process, with demand more likely to be high this period if it was high last period. Specifically:

$$\Pr(D_{t+1} = Q) = \begin{array}{ll} 1 - \theta P & \text{if } D_t = Q, \quad \theta < 1, \\ 1 - P & \text{if } D_t = Q(1-\sigma). \end{array} \tag{1}$$

Equivalently, there are four possible states of the world described by demand in the current and previous period. Good states (so labelled because they are good for investment and will normally be associated with high levels of capacity) are those that experienced high demand last period (and are thus more likely to have high demand this period), whilst bad states are those which had low demand last period. Each of these states may experience high or low current demand, so the set of states may be labelled (GH,GL;BH,BL).

The capacity (in terms of potential supply) of capital K is K/v, the variable operating cost is b/unit (constant), and there is an alternative source of supply available at a constant total unit cost c/unit, where $c > b + rv$, and where r is the gross rate of interest (including depreciation). The technology thus exhibits fixed coefficients, which means that the concept of excess capacity is well–defined.

The utility must choose investment each period not knowing future demand with certainty, nor the revenues that it will be allowed to earn next period. The sequence of moves for the utility and the regulator is shown in Figure 1. The choice of investment in period t determines the capital stock in period t+1 and occurs before demand is realized in t+1. The regulator chooses a transfer of revenues to the utility, R_{t+1}, which is determined with full information about both K_{t+1} and D_{t+1}. Figure 1 defines a "basic game" which repeats in every period $t = 0, 1, \ldots, \infty$.

We assume the regulator is concerned only with consumer surplus. This assumption emphasizes the opportunity of the regulator to engage in strategic behavior with respect to sunk investments – there would be no problem if the regulator were indifferent as to the distribution of income between consumers and the utility (or its shareholders), nor if the regulator were acting primarily in the interest of the utility.[7] In that case, conditional on the choice of investment, the "one–shot" game is zero–sum and has a unique Nash equilibrium in which the regulator sets price equal to variable cost and the utility does not invest.

It is well known that the infinitely repeated game can support an infinity of Subgame Perfect Nash Equilibria (SPNE) that are Pareto–preferred to the Nash equilibrium of the one–shot game. Our interest is investigating how institutional constraints that limit the strategy choices which are available to the regulator might restrict the set of SNPE. Formally, we suppose the game includes a "legislator", who at $t = 0$ specifies the set of strategies that are legally permissible for the regulator.

---

[7] Our characterization is an extreme form of a political orientation which appears to describe recent regulatory objectives in some jurisdictions in the United States. The regulator is thus "captured" in the sense described by Stigler (1971) and Peltzman (1976), but the capture is by consumers, not producers.

Before comparing alternative regulatory regimes as defined by permissible strategy sets, consider as a benchmark the cost–minimizing investment path. Suppose that efficient investment has $K = vQ$ in good states and $K = vQ(1-\sigma)$ in bad. For this to be efficient, the net cost of supply must be minimized at these levels. In good states (superscripted g) the expected cost is

$$C^g = \underset{K}{\mathrm{Min}}\Big\{(1-\theta P)[b\mathrm{Min}\{Q,K/v\} + c\mathrm{Max}\{Q-K/v,0\}] + \theta P[b\mathrm{Min}\{Q(1-\sigma),K/v\} + c\mathrm{Max}\{Q(1-\sigma)-K/v,0\}] - rK\Big\},$$

and for the optimum choice of K to be vQ:

$$(1-\theta P)(c-b) > rv. \tag{2}$$

Likewise, in bad states (superscripted b) the expected cost is

$$C^b = \underset{K}{\mathrm{Min}}\Big\{(1-P)[b\mathrm{Min}\{Q,K/v\} + c\mathrm{Max}\{Q-K/v,0\}] + P[b\mathrm{Min}\{Q(1-\sigma), K/v\} + c\mathrm{Max}\{Q(1-\sigma)-K/v,0\}] - rK\Big\},$$

and for $K = vQ(1-\sigma)$:

$$rv > (1-P)(c-b). \tag{3}$$

For future reference it might be useful to think of values for the cost parameters, c, b and rv. The value of alternative sources of supply, c, will depend on whether small amounts are needed, in which case purchasing power from nearby utilities may be feasible at a typical figure in 1988 of c = 9 cents/KWhr, or whether substantial shortfalls must be made up, in which case the shortage costs, c, may be very high indeed. Typical values for the costs incurred by the utility might be b = 2 cents/KWhr, rv = 5 cents/KWhr for nuclear, and b = 4 cents and rv = 3 cents for coal. Thus for nuclear (taking the lower value of c), $K = vQ(1-\sigma)$ in bad states and vQ in good states requires $(1-\theta P) > 5/7 > (1-P)$, whilst for coal $(1-\theta P) > 3/5 > (1-P)$. Both will be satisfied for $\theta = P = 1/2$.

III Alternative Regulatory Regimes

Strictly interpreted, rate of return regulation implies that the regulator must transfer sufficient revenues in each state to ensure that the utility earns at least a normal rate of return on its invested capital. This policy leads to the over–capitalization problem described in Averch and Johnson (1962), Baumol and Klevorick (1970), and others. Strict rate of return regulation is not mandated by legal precedent. In its most decisive opinion on the scope of utility regulation (*FPC v. Hope Natural Gas*), the U.S. Supreme Court concluded that "rates which are adequate to attract capital for growth ... are not unreasonable", but the Court did not say that the regulators are obligated to provide at least a normal rate of return at all times. Rates that persistently fail to compensate firms for their invested capital could run foul of the due process protection of the Constitution (as expressed, for example, in *Smythe v. Ames*, 1898). This difficulty does not arise here because we impose the individual rationality condition that in a SPNE the value of the firm must be non–negative, otherwise the utility would choose to abandon its franchise.

In the postwar history of utility regulation in the United States, utilities frequently earned rates of return that exceeded their capital costs (see Joskow, 1974).[8] However, in recent years, regulators have disallowed compensation for investments that were considered to be either "not used or useful" or "imprudent". Typically, the first case has involved expenditures on discontinued or unnecessary plants and the second expenditures on nuclear plants that have experienced large cost overruns. In either case, the net effect is to reduce actual earned rates of return below a normal return on investment in special circumstances.

---

[8] Greenwald (1984) discusses the view that a fair rate of return might be one which equated the market to book value of the stock. In the 1960s the ratio of the market to book value was greater than 2.

In this model, the restriction on strategies is formally interpreted to require at least a normal return on capital except in circumstances in which there is excess capacity. The consequences of this restriction are explored below.

III.1 "Used and Useful" Rate of Return Regulation

We define used and useful rate of return regulation (UUROR) as requiring the regulator to pay at least a normal return on invested capital up to an amount equal, in our model, to realized demand, but the regulator has the latitude to offer rates that fail to compensate the utility for investment when there is excess capacity. Suppose that the regulator pays a gross return r on invested capital up to $vQ(1-\sigma)$, and thereafter $s(i)$ in the demand state $i = \{H,L\}$ on the additional capital, up to a maximum of vQ. This policy is described by the revenue function

$$R(i, K_t, Q_t) = bQ_t + r.\mathrm{Min}\{K_t, vQ(1-\sigma)\} + s(i).\mathrm{Min}\{vQ, \mathrm{Max}[K_t - vQ(1-\sigma), 0]\}.$$

Let $s(H) = s$ and $s(L) = s'$. A policy with $s \geq r$ and $s' \leq r$ is permissible under UUROR. The expected profit to the utility in good states is

$$\pi^g = \left\{(1-\theta P)s + \theta Ps' - r\right\}\{K - vQ(1-\sigma)\}, \quad vQ(1-\sigma) \leq K \leq vQ, \tag{4}$$

and to induce the utility to set $K = vQ$, $\{(1-\theta P)s + \theta Ps' - r\} \geq 0$. Expected profit in bad states is

$$\pi^b = \left\{(1-P)s + Ps' - r\right\}\{K - vQ(1-\sigma)\}, \quad vQ(1-\sigma) \leq K \leq vQ, \tag{5}$$

and to induce the utility to install $K = vQ(1-\sigma)$, $\{(1-P)s + Ps' - r\} < 0$. Let $s = r + \rho$ and $s' = r - \mu$, with $\rho, \mu > 0$ to satisfy the efficiency constraints. The cost minimizing choice of ρ, μ is to set $\pi^g = 0$, or

$$\rho = \theta P\mu/(1-\theta P). \tag{6}$$

Provided $\rho, \mu > 0$, $\pi^b = 0$ as $K^b = vQ(1-\sigma)$, so the utility will earn zero expected profits in both states.

The next step is to check that the regulator would not wish to have $K = vQ$ in all states, which it could achieve by setting $(s, s') = (r+\varepsilon, r+\varepsilon)$, ($\varepsilon$ small), nor $K = vQ(1-\sigma)$ in all states, which it could achieve by setting $(s, s') = (r, r-\varepsilon)$. It can readily be established that for neither of these to be preferable

$$1-\theta P > \frac{rv}{c-b} > 1-P, \tag{7}$$

which is the same condition as (2) and (3). Thus the socially efficient pattern of investment would be chosen provided the regulator can commit to this policy.

Because the policy $R(\cdot)$ induces efficient investment behavior and, for values of ρ, μ satisfying (6), extracts all the consumer surplus, it is a first best policy for the regulator, again, provided that the regulator can commit herself to this policy. The desirability of this policy is summarized in

Proposition 1. A commitment to the cost–minimizing rate of return policy $R(\cdot)$ is a first–best policy for the regulator.

The only difficulty that $R(\cdot)$ might pose for the regulator is that there need not exist an (s, s') policy which supports the efficient investment plan as a SPNE of the repeated game. To investigate this question, we need a more formal specification of the repeated game without commitment.

IV The sustainability of regulation

If the regulator cannot be bound to her announced policy (of setting an agreed rate of return or price), after the utility has invested, then we need to inquire whether the regulator will ever have an incentive to deviate from her stated policy and whether she can credibly commit to not deviating.[9] We can model this problem as a game played between the regulator and the utility. In each period the regulator can either continue with the policy, or she can deviate. After observing the regulator's choice, the utility can then either continue investing as though the regulator were committed to continue her policy into the future, or, if the regulator has deviated, it can adopt a different investment plan. We can then investigate the set of perfect Nash equilibria in this repeated game.

Formally, the game can be described in much the same way as the model of collusion under uncertainty in Green and Porter (1984). The regulation game is an infinite repetition of a "basic game", the moves of which are summarized in Figure 1. The information when the utility invests is realized demand in period t–1, which signals whether the state is good or bad. In addition, the utility looks ahead to the return it can anticipate on its invested capital. After demand is realized in period t, the regulator chooses a revenue transfer function $R_t(\cdot)$, conditional on the realized demand, and the utility then chooses output, $Q_t \leq \min\{K_t/v, D_t\}$. The payoff to the utility at the termination of the basic game is $\pi_t = R_t - bQ_t - rK_t$, and the payoff to the regulator (relative to the alternative of purchasing Q_t at a cost, c) is

---

[9] The issue of credibility has been discussed by Kydland and Prescott (1977) in the macroeconomic context, and by Roberts (1983), who looks at limitations on the taxation of capital (though in a setting of asymmetric information, in contrast to our model).

$U_t = cQ_t - R_t$ (provided the regulator acts within the constraints imposed by the legislator). Both the utility and the regulator are risk neutral.

The unique Nash equilibrium of the basic game is $\{R, K\} = \{bQ_t, 0\}$. The regulator will price at variable cost, b, and the utility does not invest (and therefore $Q_t = 0$). Any investment would be expropriated by the regulator, who will price at variable cost in order to utilize available capacity. Given this behavior by the regulator, there is no reason for the utility to invest.

The regulation game is an infinite repetition of the basic game. The regulator chooses a revenue function $\{R_t(\cdot)\}$ to maximize $E[\Sigma_{t=0}^{t=\infty} \beta^t U_t(R_t,Q_t)]$, where β is the discount factor. The utility chooses $\{Q_t, K_{t+1}\}$ to maximize $E[\Sigma_{t=0}^{t=\infty} \beta^t \pi_t(R_t,Q_t,K_t)]$, subject to $Q_t \leq \min\{K_t/v, D_t\}$. The regulator and utility have complete information on the history of demand and actions by each party. Note that as a consequence of the sequential nature of the basic game, the information available to each party in period t is different. The regulator knows K_t and D_t, while the utility knows only D_{t-1} before it chooses K_t, and it knows both D_t and $R_t(\cdot)$ when it chooses Q_t.

A *contingent strategy* for the regulator is an infinite sequence $S^r = \{S_0^r, S_1^r, \ldots, S_t^r, \ldots\}$, where S_t^r determines a revenue function as a function of history, $S_t^r(h_t^r) = R_t(\cdot)$. Similarly, a contingent strategy for the utility is an infinite sequence $S^u = \{S_0^u, S_1^u, \ldots, S_t^u, \ldots\}$, where K_0 is given, and S_t^u determines output and next period's capital as a function of the history, h_t^u, available to the utility: $S_t^u(h_t^u) = (Q_t, K_{t+1})$.

A strategy profile is defined as the pair of contingent strategies $\{S^r, S^u\}$, and expectations are taken with respect to the probability distribution of prices and outputs this induces. A Nash equilibrium is then a strategy profile $\{S^{r*}_, S^{u*}\}$ which satisfies

$$\underset{S^{r*}_, S^u}{E}[\Sigma_{t=0}^{t=\infty} \beta^t \pi_t\{S_t^u(h_t^u), S_t^{r*}(h_t^r)\}] \leq \underset{S^{r*}_, S^{u*}}{E}[\Sigma_{t=0}^{t=\infty} \beta^t \pi_t\{S_t^{u*}(h_t^u), S_t^{r*}(h_t^r)\}],$$

and

$$\underset{S^r, S^{u*}}{E}[\Sigma_{t=0}^{t=\infty} \beta^t U_t\{S_t^{u*}(h_t^u), S_t^r(h_t^r)\}] \leq \underset{S^{r*}, S^{u*}}{E}[\Sigma_{t=0}^{t=\infty} \beta^t U_t\{S_t^{u*}(h_t^u), S_t^{r*}(h_t^r)\}],$$

for all feasible strategies S^r, S^u. In all cases we assume that the utility acts as a Nash follower taking the actions of the regulator as given.

We confine our investigation of equilibrium strategies to those that are subgame perfect. Equilibria will differ according to the the legislative constraints which restrict the regulator's choice of revenue transfers, and the feasible responses to deviations by the participants. In game 1 below, we assume that if the regulator does not follow a UUROR revenue function, then the game will degenerate into the one–shot Prisoner's Dilemma. Because the utility acts as a Nash follower, the only rational deviation by the regulator is complete expropriation of the sunk capital. This action could be interpreted as not legally permissible under the rules of UUROR. Therefore in game 2 we modify game 1 to remove the expropriation option.

We can now inquire whether rate of return regulation is sustainable for the case in which the efficient equilibrium requires variations in capacity – that is, equations (2) and (3) are satisfied. The game is specified as

Game 1 Rate of Return Regulation

Rate of return regulation is defined by the revenue function, defined as:

$$R(i, K_t, Q_t) = bQ_t + r.\mathrm{Min}\{K_t, vQ(1-\sigma)\} + s(i).\mathrm{Min}\{vQ, \mathrm{Max}[K_t - vQ(1-\sigma), 0]\}. \tag{8}$$

As before, the regulator promises to pay a gross return r on capital installed up to $vQ(1-\sigma)$, and thereafter $s(H) = s = r + \rho$, $s(L) = s' = r - \mu$ on capital up to a maximum of vQ. The values of ρ, μ are chosen to satisfy (6), and are thus cost minimizing. The regulator's strategy is $S_t^r(h_t^r) = R_t$:

$$R_0 = R(i_0,K_0,D_0),$$

$$R_t = R(i_t,K_t,D_t), \quad \text{if } K_t \geq vQ(1-\sigma) \text{ and } R_{t-1} \geq R(i_{t-1},K_{t-1},D_{t-1}),\ t > 0, \tag{9}$$

$$= bQ_t, \quad \text{otherwise.} \tag{10}$$

The utility's strategy is the choice of capital and output that maximizes its present discounted profit, taking S_t^r as given. Of course, the utility will only invest if the expected present discounted value of the franchise, V^u or W^u, is non–negative.

We can now examine the perfect equilibria of this game. The one–shot "Prisoner's Dilemma" outcome $\{bQ_t; 0\}$ is clearly an equilibrium, for if $R_t = bQ_t$, $K_{t+1} = 0$, and then by (10), $R_{t+1} = bQ_{t+1}$ and so on. The next question is whether the efficient solution $\{(r; s, s'); (vQ, vQ(1-\sigma))\}$ is also an equilibrium (adopting the natural shorthand that the first bracket characterises the form of regulation, the second the level of capital in the two states). This in turn involves checking that the regulator would not wish to deviate from (9), given that it would lead to a SPNE with zero investment next period.

The advantage to the regulator of deviating is the value of the current capital stock, the return on which is effectively expropriated. The disadvantage is the cost of having to meet future demands from other sources, or the cost of any shortages which might result. In the present framework the regulator may wish to deviate in state GH or GL, depending on the size of ρ, μ. This case is analyzed in the Appendix. The simpler case to examine is when the utility underinvests, so that $K_t = K \leq vQ(1-\sigma)$, which is a possible equilibrium of the game. (If $\sigma \to 0$, this is equivalent to the case analyzed in the Appendix). Suppose that the period under question is sufficiently long that capital completely depreciates if there is no new investment.

(The continuous time case is dealt with in the next footnote and yields essentially the same results). The value to the regulator of paying the rate of return on the full capacity, rather than having no capacity, is

$$U^r = (c-b-rv)K/v \tag{11}$$

per period. The present value of this stream of benefits is $V^r = U^r/(1-\beta)$. The advantage of deviating is the one period capital cost saving, rK, and the cost is βV^r from losing the benefits of capital from next period. Deviating is unattractive if

$$rK < \beta V^r = \frac{\beta}{1-\beta}(c-b-rv)K/v,$$

or

$$\xi \equiv \frac{c-b}{rv} > \frac{1}{\beta}. \tag{12}$$

If this condition is satisfied, the regulator will have no incentive to expropriate the utility's capital stock when the utility maintains a constant level of capacity, regardless of its level.[10] If

---

[10] In continuous time the problem can be set up as follows. Let $dK/dt = I - \delta K$, where I is gross investment, and δ is the rate of depreciation. If demand grows at rate g, and the rate of interest is i (so that $r = i + \delta$ is the gross rate of return), the present discounted cost of supplying electricity under rate of return regulation is

$$\int_0^\infty (b+rv)Q_0 e^{gt} e^{-it} dt = (b+rv)Q_0/(i-g).$$

If the regulator deviates the utility supplies $Q_0 e^{-dt}$ at date t,the balance being met at shortage cost c. The present discounted cost is

$$Q_0 \int_0^\infty \{be^{-dt} + (e^{gt} - e^{-dt})c\} e^{-it} dt = Q_0\{c/(i-g) - (b-c)/(i+\delta)\}.$$

Deviating is unattractive if the second cost exceeds the first; that is, if

$$\frac{c-b}{rv} > \frac{i+\delta}{\delta+g} = \frac{r}{\delta+g}.$$

the utility increases capital to vQ in good states, the regulator will be tempted to expropriate unless $\xi > \xi^*$, where the derivation of ξ^* is discussed in the Appendix.

There are two possibilities. If $\xi > \xi^*$, the regulator will have no incentive to expropriate the utility's capital stock in any state, and the efficient equilibrium $\{(r;\ s,\ s');\ (vQ,\ vQ(1-\sigma)\}$ will be sustainable as a SPNE. If $\xi^* > \xi > 1/\beta$, the regulator would expropriate if $K = vQ$, but would not expropriate if $K \leq K^* < vQ(1-\sigma)$. The utility, knowing this, will set $K_H = K^*$, and the equilibrium will be $\{(r;\ s,\ s'),\ (K^*, vQ(1-\sigma))\}$, which is inefficient.

If $P = 0.85$, $\theta = 0.5$, $\sigma = 0.2$, $\beta = 0.2$, $rv = 5$, $b = 2$ cents/KWhr, then $\xi^* = 5.76$, and $\xi > \xi^*$ if $c > 31$cents/KWHr. This is not an unreasonably high figure for the cost of failing to have reliable power supplies, which is likely to be very large indeed. Another way of looking at the condition is that if $c = 31$cents/KWhr, then $\beta > 0.2$. This condition may be interpreted in terms of the length of the period over which the capital stock depreciates. Thus, if the rate of discount is 9 per cent real, the length of the time period corresponding to $\beta = 0.2$ is 18.6 years. (Properly speaking, this period should be compared with the half–life of the capital stock, and as such is high). Higher shortage costs, c, will make deviations less attractive.

V Trigger strategies for small deviations

As we remarked earlier, the rules of UUROR are likely to make it illegal to expropriate the entire capital stock, though they do not restrict the regulator from paying a low, possibly zero, return on excess capacity. Instead of expropriating the whole capital, therefore, the regulator is effectively limited to expropriating surplus capital. Knowing this, the utility's best response to such a minor deviation is to avoid any excess capacity by underinvesting. Of course, if the utility does not satisfy the conditions of the franchise and supply at least base load capacity, then the regulator is no longer bound by the rules of UUROR, and may

---

The worst case is when $g = 0$, and if $r = 12\%$, $\delta = 3\%$, then the condition is $(c-b)/rv > 4$, similar to the condition in the text, and guaranteed by $c > 22$ cents.

expropriate in such cases.

Formally, we can set this up as the following

Game 2 Threat-constrained rate of return regulation

The game is essentially the same as game 1 except that the regulator cannot expropriate except in response to a failure to supply base load capacity. Rate of return regulation is defined by the same revenue function as in game 1, equation (8) with $s(H) = s = r + \rho$, $s(L) = s' = r - \mu$, and ρ, μ satisfying (6). The game is "threat–constrained" in the following respect. If the utility fails to provide the efficient level of capacity when demand is high, but invests enough to meet base load demand, the regulator will pay a price $p_t = b+rv$ in the next period (i.e. will set $s' = 0$ and pay no return on unutilized capacity). This will be sufficient for the utility to break even on its installed capacity. If the utility fails to meet base load, the regulator will pay b and the market will collapse to the one–shot Prisoner's Dilemma equilibrium of the basic game. Thus the regulator tolerates underinvestment which meets base load, but reverts to the myopic strategy if the utility fails to perform to this minimum level.

The regulator's strategy is $S_t^r(h_t^r) = R_t$:

$$
\begin{array}{lll}
R_0 & = R(i_0, K_0, H) & \\
R_t & = R(i_t, K_t, D_t) & \text{if } K_t \geq vD_{t-1} \text{ and} \\
 & & \quad R_{t-1} \geq R(i_{t-1}, K_{t-1}, D_{t-1}),\ t > 0, \\
 & = (b+rv)Q_t & \text{if } vQ(1-\sigma) \leq K_t < vD_{t-1},\ t > 0, \\
 & & \text{or } (b+rv)Q_\tau \leq R_\tau < R(i_\tau, K_\tau, D_t) \text{ for some } \tau,\ 0 \leq \tau < t, \\
 & = bQ_t & \text{if } K_t < vQ(1-\sigma).
\end{array}
$$

Now consider what would happen if the regulator reduced the revenue in state GL from $(b+rv)Q(1-\sigma) + \sigma s'vQ$ to $(b+rv)Q(1-\sigma)$. Given these strategies the utility would predict that all future prices would be either b or b+rv. At prices b+rv it would be unattractive to increase capacity above $vQ(1-\sigma)$ in good states, and it would be attractive to continue investing at least $vQ(1-\sigma)$ at this predicted price. Thus b+rv is a rational forecast if the utility invests $vQ(1-\sigma)$.

The regulator can deviate from $R(\cdot)$ by moving to $(b+rv)Q_t$. (It would not be rational to choose any other compensation if the regulator deviates from R(.)). If she sets price at $p^* = b+rv$ in state GL the one–period gain is $(r-\mu)vQ\sigma$, and the loss is the expected present value of having capacity $vQ(1-\sigma)$ at price p^* rather than $\{(r,s,s'); (vQ(1-\sigma),vQ)\}$. The expected gain to the regulator of having rate of return regulation in good states rather than $vQ(1-\sigma)$ at $p^* = b+rv$ is

$$U^g/Q = (1-\theta P)\{c-b-sv)\sigma - \theta P\sigma s'v = \{(1-\theta P)(c-b) - rv\}\sigma, \tag{14}$$

whilst the gain in bad states is zero. The present value of the (r; s, s') regime to the regulator starting in a bad state is therefore

$$W^r = \frac{\beta(1-P)U^g}{(1-\beta)\{1-\beta P(1-\theta)\}}, \tag{15}$$

(where the algebra is presented in the Appendix). The one period gain from deviating will be unattractive if

$$(r-\mu)vQ\sigma < \beta W^r. \tag{16}$$

Next, consider the consequences of reducing the revenue in state GH from $(b+rv)Q + \sigma\rho vQ$ to $(b+rv)Q$ – i.e. not paying the premium on capital of ρ which was designed to compensate for the shortfall of μ in the rate of return in low demand states. The utility will respond to this minor deviation by setting $K = vQ(1-\sigma)$. The one period gain from the deviation is $\sigma\rho vQ$, and the cost is βV^r (where V^r is the value to the regulator of continuing with efficient investment under $R(\cdot)$, rather than moving to the underinvestment equilibrium, starting from a good state). The deviation will be deterred if

(17) $$\sigma\rho vQ < \beta V^r.$$

The regulator must now choose the level of ρ (or μ) to satisfy constraints (16) and (17), if possible. Deviation in state GL can be made less likely by increasing μ, but, by (6) this requires that ρ be increased, and thus increases the likelihood that deviation in state GH will become attractive. The borderline case will occur when the only credible UUROR which supports the efficient investment plan increases ρ, μ to the point where the incentive to deviate is equal in states GH, GL:

(18) $$\frac{r-\mu}{W^r} = \frac{\rho}{V^r},$$

which, from the formulas given in the Appendix, can be written

$$\beta(1-P)\rho = (1-\beta P)(r-\mu).$$

Together with the cost–minimizing relationship between ρ and μ given in (6), this can be solved to give

(19) $$\mu = \frac{r(1-\beta P)(1-\theta P)}{1-P(\theta+\beta-\theta\beta)}; \quad \rho = \frac{(1-\beta P)\theta Pr}{1-P(\theta+\beta-\theta\beta)}.$$

If ρ, μ are set at these levels, deviations will be deterred provided (16) (or, equivalently, (17)) is just satisfied, i.e. provided ξ exceeds some critical value $\hat{\xi}$, say.

The various possible equilibria of the game can now be described. If $\xi \geq \hat{\xi}$, then the outcome will be the efficient equilibrium $\{(r;\ s,\ s');\ (vQ,\ vQ(1-\sigma)\}$. If $\hat{\xi} > \xi \geq 1/\beta$, the utility will set $K = K^* < vQ(1-\sigma)$, and the outcome will be the underinvestment equilibrium $\{(r;\ s,\ s');\ (K^*,\ vQ(1-\sigma)\}$. Finally, if $\xi \leq 1/\beta$, then the best the regulator can do is set $p = p^* = b+rv$, so $\{b+rv;\ vQ(1-\sigma)\}$ is a perfect equilibrium. Summarizing:

Proposition 2 Under "used and useful" rate of return regulation, it may be desirable to pay a below normal rate of return on excess capacity, which will be compensated by above normal rates of return at other times, in order to sustain the efficient pattern of investment. Even then, it may not be possible to sustain the efficient equilibrium and the utility may underinvest.

It is worth stressing that the flexibility provided by the "used and useful" doctrine has advantages and drawbacks. On the one hand the ability to reduce s' and raise s allows the regulator to transfer profits between states GL, GH, and thus to reduce the incentive to deviate (by either setting $s' = 0$ or $\rho = 0$). This may be sufficient to deter such deviations, and thus allow the efficient equilibrium to be sustained. On the other hand the freedom to pay less than the normal return in low demand states provides the regulator with the opportunity to exploit the utility. Whether the advantage outweighs the disadvantage turns on the response of the utility to small deviations, and the likelihood that the players will make costly errors, neither of which is adequately modelled in our stylized game.

VI Price Regulation

The main conclusion of rate of return regulation was that it is efficient, provided that it is sustainable, but that it might not be sustainable within the legislative framework we examined. This raise the question as to whether there is an alternative form of regulation which may be superior to UUROR. Britain, with its very different institutional history and regulatory legislation, has recently privatized a number of previously nationalised utilities – gas, telecoms – and is about to privatize the electric supply industry. In each case the legislated method of regulation has been to limit the price that the utility can charge on its regulated sales, not to set the rate of return. Typically, the utility is constrained from setting a price higher than a figure determined by an annually adjusted index formula. This formula may include fuel costs, the consumer price index, or other observable price indices, but usually excludes cost data from the regulated utility. The advantages of this approach are clear – the price ceiling provides incentives for cost minimization and rewards improvements in efficiency in a way absent from cost–based methods of regulation. Our model does not address these incentive issues, but we can ask whether these advantages come at a cost in terms of the issues we do discuss. Can price regulation do as well as rate of return regulation by the standards we discuss – efficiency, measured by the choice of investment, cost to consumers, and sustainability?

Price regulation can take almost any form, but there are a number of specifications that seem natural in the present setting. The distinctive feature of price regulation is that the utility receives no return on unutilized capacity, though this may not be as significant a difference as might at first appear. If we initially restrict attention to *linear* price schedules, then it is natural to consider specifying four state–specific prices, p_i, $i \in \{GH,GL;BH,BL\}$. In good states the utility will earn expected profits

$$(23) \qquad \pi^g = (1-\theta P)(p_{GH}-b)K/v + \theta P(p_{GL}-b)\mathrm{Min}\left\{Q(1-\sigma),\ K/v\right\} - rK, \quad K \le vQ,$$

and the condition for choosing $K = vQ$ is

$$(24) \qquad p_{GH} \ge b + rv/(1-\theta P).$$

The regulator will ensure this is satisfied with equality, and can set $p_{GL} = b$, at which level expected profits, π^g, are held to zero. In the bad state the utility will earn expected profits

$$(25) \qquad \pi^b = (1-P)(p_{BH}-b)K/v + P(p_{BL}-b)\mathrm{Min}\left\{Q(1-\sigma), K/v\right\} - rK, \;\; K \geq vQ(1-\sigma).$$

There are a wide range of prices satisfying the condition that the utility will choose $K = vQ(1-\sigma)$ and make non–negative profits[11] — provided $p_{BH} < b+rv/(1-P)$. The natural choice is $p_{BL} = p_{BH} = b+rv$, which again minimizes the expected consumer cost by holding profits to zero.

The ex–post realized profits will be

$$\pi_{GH} = \frac{rvQ\theta P}{1-\theta P}; \quad \pi_{GL} = -rvQ,$$

and zero in the other two states. (The convention on notation here is that π^g is the expected profit in good states, π_{GH} is the realized profit in state GH). Compared to rate of return regulation, this form of price regulation has the same expected cost but greater variability from period to period unless $\rho > r\theta P/\{\sigma(1-\theta P)\}$.

If the regulator were to choose a single price high enough to induce the efficient level of investment, then the price would have to be $b+rv/(1-\theta P)$, and the utility would now make expected profits in good and bad states, at the expense of the consumers. If the regulator were to choose prices contingent only on current demand (high or low), then again a price $p_H = b+rv/(1-\theta P)$ is required to induce efficient investment. This will insure that the utility makes positive expected profits in good states, and, as a result, the price in low demand states, p_L, can be held below average cost, b+rv, resulting in expected losses in bad states. The only constraint is that the present value of profits, starting from either good or bad states, be non–negative. If $\pi^g > 0 > \pi^b$, then the formulas in the Appendix show that $V^r > W^r$, so the binding constraint will be $W^r = 0$, and the utility will make strictly positive expected profits in good states, at the cost of higher prices for consumers than rate of return regulation.

---

[11] If the utility is to be held down to zero expected profits in good states, it must be ensured non–negative expected profits in bad states if the present value of holding the franchise is to be non–negative.

If the regulator can specify a non–linear price schedule, the two–part tariff described by the revenue function

$$R(q_t) = bQ_t + rv.\mathrm{Min}\{Q_t, vQ(1-\sigma)\} + \frac{rv}{1-\theta P}.\mathrm{Max}\{0, Q_t - vQ(1-\sigma)\} \qquad (26)$$

is the natural analog of the UUROR (r,s,s').

This pays b+rv on the first vQ(1–σ) units, and b+rv/(1–θP) on additional sales, and will induce the efficient level of investment. Like the four–state price rule, it yields zero expected profit in good and bad states. The ex–post realized profits will be

$$\pi_{GH} = \frac{\sigma rvQ\theta P}{1-\theta P}; \quad \pi_{GL} = -\sigma rvQ,$$

and zero in the other two states. The variability of profits will thus be lower (only σ times the size) of those in the four–state system of price regulation.

Thus if prices are either non–linear or state–contingent, they can induce efficient investment at least cost to consumers as effectively as rate of return regulation. The remaining question is whether there is a system of price regulation which is as good as or better than rate of return regulation as far as sustainability is concerned. The major problem with sustaining the efficient outcome is that the short–term benefits of deviating may be too great. The temptations to deviate are higher, the more profits of the utility can be appropriated by the regulator. Under UUROR, we saw that the temptation to deviate could be minimized by raising ρ, μ, and thus reducing utility profits in GH whilst reducing losses in GL, until the incentives to deviate were equalized in both states. Under price regulation, if the regulator can make state–contingent transfers T_i in state i, then it would be possible to exactly replicate UUROR. In fact, it is also possible to do better than UUROR, because there are now four instruments, T_i, rather than two, ρ and μ. If all four transfers are non–zero, though, the cost to consumers will be higher than under UUROR. Consider the set of transfers:

$$(27)\qquad T_{GH} = -A - \theta PB;\quad T_{GL} = -A - (1-\theta P)B;$$

$$T_{BH} = C - DP;\quad T_{BL} = C + (1-P)D,$$

where A, B, C ≥ 0. The expected profits are now $\pi^g = -A$, $\pi^b = C$, and for the utility to wish to continue investing, the present discounted value of profits, V^u, $W^u \geq 0$. If A > 0, then $V^u = 0$ implies $W^u \geq 0$, and for $V^u = 0$,

$$A = \frac{\beta\theta PC}{1-\beta P}.$$

Then A, B, and D can be chosen to make the gains from minor deviations equal in each of the four states, rather than just equating them in two states, as with UUROR. These two conditions, together with two boundary conditions at which the incentives to deviate are exactly balanced by the costs (the four state counterparts to (16) and (17) holding with equality) give four equations in A, B, D, and $\xi \equiv (c-b)/rv$. Since A, B, and D > 0, the condition on ξ is strictly weaker than under UUROR. Formally, we can model unconstrained price regulation as

Game 3 Threat-constrained non–linear price regulation

Price regulation is defined by the revenue function given in (26) and by state–contingent transfers $T_t = T(i_t)$ defined by (27). We continue to suppose that as in Game

2 the regulator cannot expropriate the utility except in response to failing to supply base load. The regulator's strategy is $S_t^r(h_t^r) = R_t + T_t$:

$$
\begin{aligned}
S_0^r &= R(Q_0) + T_0, && \\
S_t^r &= R(Q_t) + T_t && \text{if } K_t \geq vD_{t-1}, \text{ and } R_{t-1} \geq R(Q_{t-1}), \ t \geq 0 \\
&= (b+rv)Q_t && \text{if } vQ(1-\sigma) \leq K_t < vD_{t-1}, \ t > 0, \\
& && \text{or } (b+rv)Q_\tau \leq R_\tau < R(Q_\tau) \text{ for some } \tau, \ 0 \leq \tau < t, \\
&= bQ_t && \text{if } K_t < vQ(1-\sigma).
\end{aligned}
$$

It will be seen that if $A = C = D = 0$, and $B = s'vQ\sigma/(1-\theta P)$, then the transfers to the utiliity are identical to those of Game 2, and in this case Game 3 will have the same equilibria. As argued above, A, C, and D can be increased above zero, thereby increasing the range of values of ξ that are sustainable as efficient SPNE. However, if $A > 0$, the present discounted value of utility profits in bad states, W^u will be strictly positive, implying that the present discounted cost of electricity will be higher than under UUROR. This extra cost is the price that must be paid to sustain the efficient equilibrium, rather than accepting the underinvestment equilibrium. Summarizing

Proposition 3. Provided the regulator can make state–contingent transfers to the utility, non–linear price regulation can replicate the efficiency and sustainability properties of rate of return regulation, and can also sustain efficient equilibria which could not be sustained under rate of return regulation, though at a higher cost to the consumer.

It should be clear that the state contingent transfers are necessary, not only to improve upon UUROR, but even just to achieve the same degree of sustainability. Whereas such transfers are very natural (and indeed are the essence of) UUROR, they are far from natural in the context of price regulation. Because rate of return regulation pays a (possibly state–contingent) return on the total capital, it has a natural base on which to make the required state–contingent transfers. Price regulation specifically excludes payment on unutilized capacity, and thus lacks a natural mechanism for making these transfers. Thus, whilst in our model it is possible to improve upon "used and useful" rate of return regulation, it is not clear how well such a system of regulation would function in practice. Our model assumes full information, no informational assymetries nor any moral hazard, and also that both parties agree on the transition probabilities θ, and P. If these conditions were not satisfied, then it may yet turn out that rate of return regulation were more robust than the potentially superior but more delicate state–contingent price regulation.

VI Inefficient choice of techniques and the reverse Averch–Johnson Effect

The sustainability of rate of return regulation against shortfalls in demand depends on the size of (c–b)/rv, or the ratio of current cost savings to the capital costs. It is quite possible for $c > d+rw > b+rv$, where (d, w) = (variable cost, capital per unit output) describes an alternative less efficient, but also less capital–intensive technology with $w < v$.[12] (Gas turbines rather than baseload thermal, for example). It is also possible that $(c-d)/rw > (c-b)/rv$, and that (b, v) is not sustainable whilst (d, w) is. In that case the inefficiency might take the form of selecting an insufficiently capital–intensive technolgy rather than investing in too little capacity.

---

[12] If (b, v) is to be efficient for expansionary investments in good states, compared to (d, w), then $(d-b)(1-\sigma\theta P) > r(v-w)$. Thus if $b = 2$, $rv = 5$, $d = 5$, $rw = 3$, $\sigma\theta P < 1/3$.

The standard Averch–Johnson effect postulated that utilities subject to rate of return regulation and paid a rate of return above the cost of capital would have an incentive to select excessively capital intensive technology or to overinvest. In the present case this temptation is balanced by the fear that excessive capital intensity will induce the regulator to lower the allowed rate of return and cause the utility losses. It would be pleasing if these two counteracting forces balanced, leading to socially efficient choices of capital intensity, but there are no strong reasons to believe this will happen. Rather, in periods of rapid and stable demand expansion, the standard A–J effect is likely to dominate, whilst in periods of stagnant and unstable demand, the reverse effect may well predominate. This is at least a plausible interpretation of the pre– and post oil shock periods of regulation.

VII Conclusions

Has the implicit regulatory contract failed? The tightness of regulatory constraint changed significantly in the past two decades relative to the period of sustained growth that occurred in the years after WWII (see Joskow, 1974). These events could be consistent with either of the equilibrium outcomes discussed in Sections III and IV. We showed that an efficient investment path could be sustained as a UUROR equilibrium, with high returns in "good" states and low returns in "bad". One could interpret the early postwar period as a sustained "good" state, followed by a more recent prolonged "bad" state. Alternatively, one could apply the model in Section IV and conclude that there has been a breakdown of the efficient contract, and perhaps a movement to an underinvestment equilibrium.

A survey of planned electric powerplant investments in the past two decades suggests that there has been a movement toward underinvestment. Excluding hydro and geothermal projects, which are in limited supply, in 1977, electric power utilities had more than 650 projects in the planning stage, of which about one–half were more than 500 megawatts in size. In 1986, the total number was 64, of which 7 were more than 500 megawatts. Much of this ten–fold reduction is the consequence of excess capacity resulting from the exuberant construction program of the 1960s and early 1970s. However, the sharp change in capital intensity suggests that utilities are following a risk–minimizing approach of building for a low level of demand and relying on smaller plants with shorter leadtimes in the event of greater than anticipated demand growth. This pattern of investment, along with regulatory behavior that has granted rate increases with reluctance, with liberal use of "prudence review", suggests a regulatory environment that is not inconsistent with the "weak deviation" scenario described in Section IV. Regulated prices are adequate to compensate utilities for their investments in the event that demand is adequate to assure a high level of capacity utilitization. Utilities can make this situation more likely by underinvesting in base load facilities, and relying on relatively risk–free investment in small scale facilities in the event that more capacity is needed.

The doctrine of "used and useful" allows regulators to withhold payment on capital that does not provide direct benefits to ratepayers, and this could be applied to utilities with large amounts of excess capacity or to cancelled plants. When expenditures can be excluded from the rate base, rate of return regulation assumes characteristic of price regulation, where revenues are paid only on capital that is employed in production. We show that the flexibility to engage in "used and useful" rate of return regulation has mixed benefits. It is necessary to avoid the capitalization bias inherent in strict rate of return regulation. But the opportunity to withhold compensation can undermine the credibility of regulation and encourage undervestment by the utility. The problem is more severe when the regulator is enjoined from expropriating capital that is actually used (but can expropriate excess capacity), because the utility can safely invest to meet only base load demand.

Both the utility and the regulator can take actions that make expropriation less likely and therefore contribute to a sustainable regulatory contract. By choosing a production technique that is less capital intensive (and has higher operating costs), the utility can reduce the gain to the regulator from expropriation and mitigate the risk to itself should expropriation occur. Of course this implies that the regulatory bias will be in the direction of too little, rather than too much, capital investment. The regulator can make expropriation of the utility's sunk capital stock less attractive by refusing to protect itself from adverse consequences of supply shortages. For example, the regulator can choose not to enter into purchase power agreements that would provide back–up power at a reasonable price in the event that the utility fails to invest. Not investing in interconnections with other grids or utilities would be an even stronger commitment to the regulatory compact. While this has the paradoxical result that facilitating the exchange of power across regulatory jurisdictions could have adverse consequences for efficient regulatory behavior, it is also an unreasonable policy, if only because some interchange of power is necessary to provide security in the event of forced outages.

This paper departs from a recent trend in the economics of regulation which emphasizes the importance of asymmetric information in a natural monopoly environment. The regulator must provide incentives for the firm to reveal its private information. If the regulator is interested in total economic surplus (regardless of its distribution), or if the regulator is "captured" as described by Stigler (1971) and Peltzman (1976) by the interests of the regulated industry, there are mechanisms such as those suggested by Loeb and Magat (1979) that are sufficient to achieve efficient revelation of information according to the regulator's objectives. If, however, the regulator's objectives are dominated by consumer interests, the elicitation of private information entails costs.

It is in such a political environment where consumer interests dominate regulatory behavior that sunk costs pose difficulties for efficient investment. Thus, whenever there are sunk costs, to the extent that there are problems of asymmetric information in the regulatory contract, the difficulties described in this paper will be present as well. Furthermore, we believe that the changing regulatory environment can only be adequately appreciated in a Markovian setting, with its differing incentives for opportunism in good and bad times. Our model, though simple, captures these essential features of the regulatory relationship.

References

Averch, H. and L.L. Johnson (1962), "Behavior of the Firm under Regulatory Constraint", American Economic Review, 52:1053–1069.

Bailey, E.E. (1973), Economic Theory of Regulatory Constraint, D.C. Heath, Lexington, Mass

Baron, D.P. and D. Besanko (1987), "Commitment and Fairness in a Regulatory Relationship", Review of Economic Studies, 54:413–46.

Baron, D.P. and R.B. Meyerson (1982), "Regulating a Monopolist with Unknown Costs", Econometrica, 50:911–930.

Baumol, W.J. and A.K. Klevorick (1970), "Input Choices and Rate–of–Return Regulation: An Overview and Discussion", Bell Journal of Economics, 1:162–190.

Besanko, D. (1985), "On the Use of Revenue Requirements Regulation Under Imperfect Information", in M.A. Crew, Analyzing the Impact of Regulatory Change in Public Utilities, Lexington Books, 39–55.

Demsetz, H. (1968), "Why Regulate Utilities?", Journal of Law and Economics, 11:55–66.

Federal Power Commission v. Hope Natural Gas Co (1944), 320 U.S. 591, 601

Gilbert, R.J. and J. Henly (forthcoming), "The Value of Rate Reform in a Competitive Electric Power Market, in Gilbert, R.J. (ed.), Developments in Energy Regulation.

Green, E.J. and R. H. Porter (1984), "Noncooperative Collusion under Imperfect Price Information", Econometrica, 52:87–100.

Greenwald, B.C. (1984), "Rate Base Selection and the Structure of Regulation", Rand Journal of Economics, 15: 85–95.

Goldberg, V. (1976), "Regulation and Administered Contracts", Bell Journal of Economics, 7:426–48.

Guesnerie, R. and J.–J Laffont, (1984), "A Complete Solution to a Class of Principal–Agent Problems with an Application to the Control of a Self–Managed Firm", Journal of Public Economics, 25:329–69.

Howe, S. (1983), "A Survey of Regulatory Treatment of Plant Cancellation Costs", Public Utilities Fortnightly, 52, March 31.

Joskow, P.L. (1973), "Pricing Policies of Regulated Firms: A Behavioral Approach", Bell Journal of Economics, 4:118–40.

Joskow, P.L. (1974), "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Regulation", Journal of Law and Economics, 17:291–327.

Kydland F.E. and E.C. Prescott (1977), "Rules Rather than Discretion: the Inconsistency of Optimal Plans", Journal of Political Economy, 85:473–91.

Laffont, J–J. and J. Tirole (1986), "Using Cost Observation to Regulate Firms", Journal of Political Economy, 94:614–41.

Loeb, M. and W.A. Magat (1979), "A Decentralized Method for Utility Regulation", Journal of Law and Economics, 22:399–404.

Peltzman, S. (1976), "Toward a More General Theory of Regulation", Journal of Law and Economics, 19:211–17.

Riordan, M.H. (1984), "On Delegating Price Authority to a Regulated Firm", Bell Journal of Economics, 15:108–15.

Roberts, K. (1984), "The Theoretical Limits to Redistribution", Review of Economic Studies, 51:177–95.

Sappington, D.E. (1983), "Optimal Regulation of a Multiproduct Monopoly with Unknown Technolgical Capabilities", Bell Journal of Economics, 14:453–63

Sappington, D.E. and D. Sibley (1986), "Regulatory Incentive Schemes Using Historic Cost Data", Bell Comunications and Research Working Paper

Selten, R., (1975), "Re–examination of the Perfectness Concept for Equilibrium Points in Extensive Games", International Journal of Game Theory, 4:25–55.

Smythe v Ames (1898), 169 U.S. 466, 546–7.

Spulber, D.F. (1986), "Bargaining and Regulation with Asymmetric Information about Demand and Supply", Journal of Economic Theory, 44:251–68.

Stigler, G.J., (1971), "The Theory of Economic Regulation", Bell Journal of Economics, 2:3–21.

Williamson, O.E. (1976), "Franchise Bidding for Natural Monopolies – In General and with respect to CATV", Bell Journal of Economics, 7:73–104.

Williamson, O.E. (1983), "Credible Commitments; Using Hostages to Support Exchange", American Economic Review, 83:519–40.

Appendix

The value functions for the two state Markov process can be computed as follows. Let U^g, U^b, be the one–period returns in the good and bad states respectively, and let V and W be the present discounted value of future returns starting in a good or bad state respectively, with the transitions as described by (1). Then

$$V = U^g + \beta(1-\theta P)V + \beta\theta PW,$$
$$W = U^b + \beta(1-P)V + \beta PW,$$

where in each case the second term is the expected present value of moving to the good state, and the third term is the expected present value of moving to the bad state. These two simultaneous equations can be solved for V and W:

$$V\psi = U^g + \frac{\beta\theta P}{1-\beta P}U^b,$$

$$W\varphi = U^b + \frac{\beta(1-P)}{1-\beta(1-\theta P)}U^g.$$

where

$$\psi = 1 - \beta(1-\theta P) - \frac{\beta^2\theta P(1-P)}{1-\beta P} = \frac{(1-\beta)\{1-\beta P(1-\theta)\}}{1-\beta P},$$

and

$$\varphi = 1 - \beta P - \frac{\beta^2\theta P(1-P)}{1-\beta(1-\theta P)} = \frac{(1-\beta)\{1-\beta P(1-\theta)\}}{1-\beta(1-\theta P)}.$$

For example, consider the value to the regulator of retaining rate of return regulation rather than expropriating all the returns to capital under UUROR.

$$U^g/Q = (c-b-rv)(1-\sigma) + \sigma\{(c-b)(1-\theta P) - rv\},$$

$$U^b/Q = (1-\sigma)(c - b - rv).$$

The one–period gain from expropriating is $rvQ + \rho\sigma vQ$ in state GH, and $rvQ - \mu\sigma vQ$ in state GL, and neither action will be worthwhile if

$$rvQ + \rho\sigma vQ \leq \beta V^r;\;\; rvQ - \mu\sigma vQ \leq W^r.$$

The borderline case will be when both conditions are satisfied with equality, which, together with $\rho = \mu\theta P/(1-\theta P)$, gives three equations in three unknowns, ρ, μ and $\xi \equiv (c-b)/rv$. Let the solution be ξ^*. The if $\xi \geq \xi^*$, $\{(r,s,s'); (vQ,vQ(1-\sigma))\}$ is a SPNE. If $\xi^* > \xi > 1/\beta$, then there is a K^*, $vQ > K^* > vQ(1-\sigma)$, such that $\{(r,s,s'); (K^*,vQ(1-\sigma))\}$ is a SPNE – the benefits of expropriating in GH will now be $rvQ(1-\sigma) + \{K^*-vQ(1-\sigma)\}(r+\rho) = \beta V^r$, which will allow a higher value of ρ, μ than vQ.

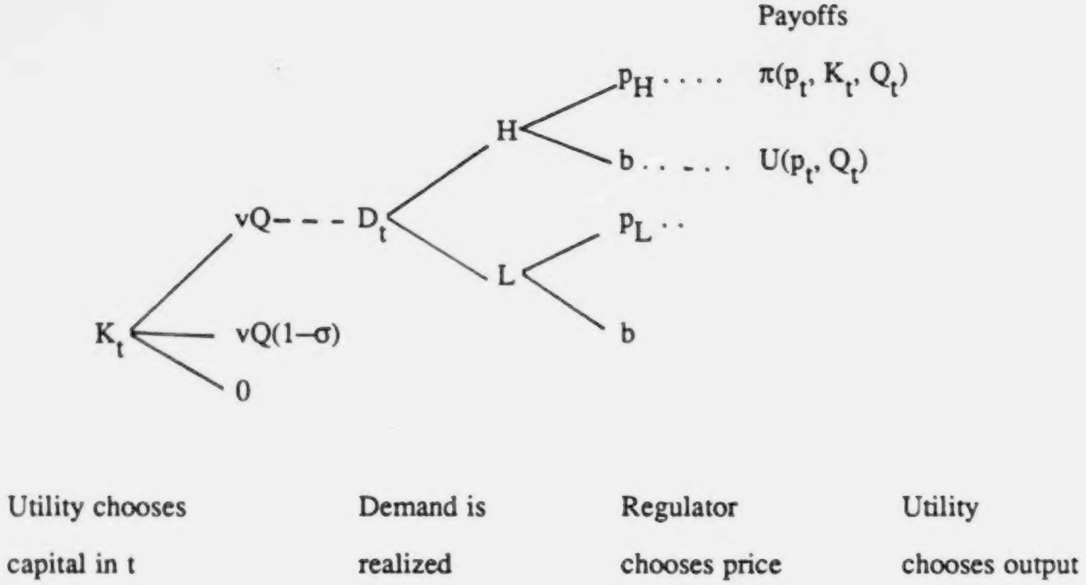

Payoffs
p_H
π(p_t, K_t, Q_t)
H
b
U(p_t, Q_t)
vQ
D_t
p_L
L
K_t
vQ(1−σ)
b
0
Utility chooses capital in t
Demand is realized
Regulator chooses price
Utility chooses output

Table 1 Basic Regulation Game

Jun-88

RECENT ISSUES OF THE WORKING PAPER SERIES OF THE DEPARTMENT OF ECONOMICS UNIVERSITY OF CALIFORNIA, BERKELEY

Copies may be obtained from the Institute of Business and Economic Research. See the inside cover for further details.

8866 Jeffrey A. Frankel
RECENT ESTIMATES OF TIME-VARIATION IN THE CONDITIONAL VARIANCE AND IN THE EXCHANGE RISK PREMIUM
Jan-88.

8867 Roger L. Ransom and Richard Sutch
CAPITALISTS WITHOUT CAPITAL: THE BURDEN OF SLAVERY AND THE IMPACT OF EMANCIPATION
Feb-88.

8868 Rudiger Dornbusch and Jeffrey Frankel
THE FLEXIBLE EXCHANGE RATE SYSTEM: EXPERIENCE AND ALTERNATIVES

Feb-88.

8869 Barry Eichengreen and Carolyn Werley
HOW THE BONDHOLDERS FARED: REALIZED RATES OF RETURN ON FOREIGN DOLLAR BONDS FLOATED IN THE 1920S
Feb-88.

8870 Roger Craine
ASSET VALUES AND ECONOMIC FUNDAMENTALS

Apr-88.

8871 David W. Roland-Holst, Sherman Robinson
THE OPPORTUNITY COST OF DEFENSE SPENDING: A GENERAL EQUILIBRIUM ANALYSIS

Apr-88.

8872 Jeffrey A. Frankel
FACTORS DETERMINING THE FLOW OF CAPITAL FROM JAPAN TO THE UNITED STATES

Apr-88.

Jun-88

RECENT ISSUES OF THE WORKING PAPER SERIES
OF THE DEPARTMENT OF ECONOMICS
UNIVERSITY OF CALIFORNIA, BERKELEY

Copies may be obtained from the Institute of Business and Economic Research. See the inside cover for further details.

8873 Roger Craine
RISKY BUSINESS: THE ALLOCATION OF CAPITAL

Jun-88.

8874 Kenneth A. Froot and Jeffrey A. Frankel
FORWARD DISCOUNT BIAS: IS IT AN EXCHANGE RISK PREMIUM?

Jun-88.

8875 Samuel Bowles and Robert Boyer
LABOR DISCIPLINE AND AGGREGATE DEMAND: A MACROECONOMIC MODEL

Jun-88.

8876 Samuel Bowles and Herbert Gintis
CONTESTED EXCHANGE: POLITICAL ECONOMY AND MODERN ECONOMIC THEORY

Jun-88.

8877 Samuel Bowles
CLASS ANALYSIS VERSUS WORLD SYSTEM? EPITAPH FOR A FALSE OPPOSITION

Jun-88.

8878 Samuel Bowles
CAPITALIST TECHNOLOGY: ENDOGENOUS CLAIM ENFORCEMENT
AND THE CHOICE OF TECHNIQUE
Jun-88.

8879 Richard J. Gilbert and David M. Newbery
REGULATION GAMES

Jun-88.

8880 Joseph Farrell and Carl Shapiro
HORIZONTAL MERGERS: AN EQUILIBRIUM ANALYSIS

Jun-88.

COLE, RAYWID & BRAVERMAN
ATTORNEYS AT LAW
SECOND FLOOR
1919 PENNSYLVANIA AVENUE, N. W.
WASHINGTON, D. C. 20006
(202) 659-9750

JOHN P. COLE, JR.
ALAN RAYWID
BURT A. BRAVERMAN
ROBERT L. JAMES
THOMAS W. FLETCHER
JOSEPH R. REIFER
FRANCES J. CHETWYND
MARGARET E. HAERING
JOHN D. SEIVER
WESLEY R. HEPPLER
PAUL GLIST
DAVID M. SILVERMAN
JAMES F. IRELAND III
MAURITA K. COLEY*
SUSAN PARADISE BAXTER
KARLENE WORTHINGTON GOLLER
ROBERT G. SCOTT, JR.*

* ADMITTED IN MICHIGAN ONLY
* ADMITTED IN VIRGINIA ONLY
* ADMITTED IN GEORGIA ONLY

CRAIG S. McCOY
(1943-1979)

CABLE ADDRESS
"CRAB"

TELECOPIER
(202) 452-0067

A PAPER DELIVERED TO THE SIXTEENTH ANNUAL
TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE

November 1, 1988

EXPLORING THE HIDDEN COSTS OF OVERBUILDS

- Paul Glist -

The recent proliferation of reports on overbuild franchises has fueled a dangerous mythology that an invisible hand will produce the best of all cable services at no public cost.

Overbuild promises are undeniably seductive. Some municipal officials are frustrated with rate deregulation and with their attendant loss of control over the extent and quality of cable service. Authorizing an overbuild would appear to offer competitive surrogates for regulation - and incidentally punish the incumbent for going over the City's head to Washington. Other cities are satisfied, but their attorneys are understandably concerned with the unsettled state of First Amendment litigation. Some federal district courts have classified cable as the electronic equivalent of a newspaper newsrack, and struck down classic RFP/franchising procedures under the Civil Rights Act.[1] Others have essentially labelled cable a natural monopoly.[2] The city's right to select from among competing franchise applicants, they have ruled, is part of its fiduciary duty to protect the public right of way.[3] Both of these views have been upheld (narrowly) by the Supreme Court, leaving cities with the unwelcome task of weighing the public interest against the comparative risks of liability.[4]

(c) Copyright 1988 Paul Glist

The cable industry's ambivalence to overbuilding is also fueled by myth. Some frustrated purchasers conclude that it is cheaper to overbuild than to buy at fair market value. Others contend that only the "bad operators" are getting "hit," and that overbuilding is a healthy shakeout of the industry.

These attitudes have developed largely because the actual facts and consequences of overbuilding have gone unreported.

In actual practice, overbuilding has not presented customers with the competitive choice from two cable systems passing their homes. For example, an overbuilder won the Fort Pierce city franchise on a promise to provide freedom of choice. Instead, it skimmed off a new development under exclusive contract and disclaimed an intention to overbuild the rest of the city.[5] It simultaneously deprives the incumbent operator of his rights to wire the entire city, and repudiates its own promise to "compete" elsewhere. In Riviera Beach, the same operator skimmed off lucrative beach side condominiums and rejected franchise provisions to serve the less lucrative inland homes.[6]

The opportunities for such creamskimming arise from a peculiarity of cable television law: access to developments has been treated as a matter of property, rather than communications law. Despite state and federal laws and rulings ostensibly opening compatible easements for cable's use,[7] many opportunities arise for de facto exclusive contracts: maintaining the privacy of rights of way and easements; reserving all cable rights in a dedication; postponing dedication until trenchs close; bulk billing; terminating easements at an MDU's outer wall or point of presence. Whatever the outcome of ongoing legal challenges to such devices,[8] their immediate result has been to frustrate open entry to developments, and to create an incentive for overbuilders to "hit" any growth market, regardless of the good quality of the existing operator.

The developer may benefit from the attendant bidding war among franchised operators but the consumer benefits are doubtful. Cable contracts are often let long before homeowners occupy dwellings, and well before homeowners gain control of their "associations." The multi-family developer plays the role of electronic censor - deciding which operator may provide the most important commodity to the fastest growing segment of the residential market. From cable's point of view, we ought to question the wisdom of recreating the RFP process with developers in control. Unlike municipalities, the Cable Act does not constrain private developers' demands for franchise fees of more than 5%, nor does it assure a renewal expectancy for a service agreement. In short, with few exceptions, overbuilding is not providing "freedom of choice," but with gerrymandered service throughout a franchise area.

The reason that creamskimming should concern franchising authorities and cable operators arises from the nature of cable television economics. Any business built on cost averaging, particularly a regulated industry which approximates universal service, is peculiarly vulnerable to creamskimming. The lessons are obvious from the other industries which have been "deregulated" in the past. One may deregulate telephone, airlines and busing, and residents in lucrative metropolitan markets will benefit from temporary price competition. But airlines have abandoned scheduled service to hundreds of more remote communities; have offered degraded service to many other sparsely populated areas; have drastically increased fares in those rural areas they continue to serve; and have neglected promptness, maintenance and safety everwhere. Bus companies have cut fares between urban markets but have abandoned intercity service to over 3,000 small communities, and have increased their rates in those rural areas they continue to serve. Telephone rates have been cut in major market long distance service, but rural phone bills have increased as much as 300% in three years. Even the rate wars caused by "deregulation" have not lasted. These "competitive" industries have begun to consolidate. After the airline rate wars of the early 1980's, mergers have created ten carriers who now control 97% of the market. In long distance telephone, the dozens of competing companies which emerged in the 1980's have failed or been absorbed.[9/] "Deregulation" of cable television can produce the same results: creamskimming which degrades or ends service to less lucrative areas; rate wars which undermine routine technical maintenance; and reconsolidation in the end.

How long will it be before economists begin to account for the erosion of the hidden rural subsidy as a public cost of deregulation? At a time when cable television should be expanding to pass all but the most remote households, creamskimming will undermine the economics which would support universal service.

One irony of overbuilding is that the most seductive promise -- the promise of rate wars -- is the most short lived and, in the long run, the most destructive. Analysts have reached general consensus that unless a market has both high housing density (e.g. 90-120 homes/mile) and high penetration, it cannot sustain two competing systems. The capital costs of erecting duplicate plants simply cannot be recovered if the pie is split in half. There may be temporary rate wars: but several analysts have reported how the technical quality of plant suffers, how maintenance is neglected and how programming funds are diverted to the war as operating revenues dry up.[10] No doubt some deep-pocket warriors are temporarily showering subscribers with upgraded customer service, free remotes, free A.O's, and "promotional" discounts, all at enormous operating losses. But I have seen less well-heeled overbuilders drive cable service in smaller markets into disrepair and ruin, with no financial way out other than buy-out or common sale to a third party. In fact, virtually all overbuilds collapse back into one franchise, with the survivor left to recapture the operating losses and redundant capital costs through rate hikes.[11] The public price of "letting the best man win" is to mortgage the future of reasonable rates.

There is another public cost to overbuilding which is largely ignored. In a true overbuild, the entire community suffers the disruption of pole changeouts, new downguys, the cutting of underground services, the mushrooming of new pedestals, trenching through lawns, gardens, driveways and fences. Every home suffers, whether the owners want a second cable or no cable at all. More and more courts and franchising authorities are desperately trying -- after the fact -- to discipline overbuilders' "negligent" cutting of incumbents' cables.[12] More and more incumbent operators are discovering that integral parts of their cable systems -- house drops, inside wiring, MDU pre-wire, as-built maps -- are being turned over to overbuilders through judicial property-law rulings which are just shy of theft.[13] None of this disruption would be welcomed as healthy "competition" by any thoughtful franchising authority; but unrestricted grants of overbuild franchises invite precisely these abuses.

Overbuild applicants are fully aware of these consequences. One must ask, then, why would a rational businessman enter a losing proposition? The answer is greenmail. An overbuilder will focus on "cream" areas, at cut rates, which presents a dilemma to the incumbent. He might "deaverage" his rates, cutting prices only in overbuilt lucrative areas. But that raises concerns that someone may prevail in ongoing challenges to "price discrimination" and "redlining." He might cut rates everwhere, in which case the loss to his existing revenues is so much greater than the gain to the overbuilder that he sets himself up for greenmail. Consider this trade press report: "Working with Guy Belello, Thomson has initiated overbuilds in Baton Rouge, Lousiana; Hammond, Louisiana, and

Mobile, Alabama; and in each case, eventually sold out to the incumbent operator."[14] Coteries of overbuilders are roaming the nation, promising "competition;" rationalizing that "it is cheaper to build than to buy;" and touting the benefits of freedom of choice. Within one to three years, however, they are cashing out at enormous profits. Does the subscriber benefit from the supposed economics of overbuilding? Or is the cable industry riding a private-market greenmail wave?

The public costs of overbuilding have been substantially unreported or under reported. Some parts of the trade press, for example, have editorially positioned overbuilders as trust-busting champions, and heralded the spate of overbuild applications as the germination of industry competition. Others, relying on clipping services, republish erroneous local press reports which fail to distinguish enabling resolutions from actual franchises. In both cases, the trade press has not "followed" the story to discover, for example, a greenmailer unmasked, who declines to accept a franchise requiring that his promises be bonded;[15] or, for another example, an overbuild collapsing into creamskimming or reconsolidation; or, for a third example, the rate increases which follow the retreat of enriched greenmailers from a market. Analysts attempting to verify press reports have found that as many as 30% of the reported "overbuilds" never happened;[16] or that rate competition never took place.[17]

One cannot contend that overbuilding is absolutely infeasible in all circumstances. But one can take to heart the 1986 recommendation of the Justice Department that the wisdom of cable television overbuilds be carefully examined on a community-by-community basis. "Generalizations about the appropriate role of overbuild competition," concluded the report, "are likely to be misleading."[18] It is absolutely appropriate for each community that is faced with an overbuild application to investigate the likely consequences of overbuild and the qualifications of the applicant. Legislation recently adopted in Florida, Tennessee, Virginia and elsewhere provides for such review, so that a community will be fully informed about economic impact, disruption, financial qualifications and other relevant criteria before leaping into overbuild.[19] Such reviews are not, as some have claimed, "pretexts" for denial. After review, some communities grant the overbuild; some deny it with record findings; some grant the application subject to anti-greenmail provisions (mandatory build-out, full construction bonds, "no sale" clauses). My view is that such review sessions are not only the most sensible way for evaluating overbuild applicants, but the most likely path to a definitive resolution of cable's "First Amendment" rights.

The legal fiction adopted by First Amendment fundamentalists is that the construction of cable systems should be given the same constitutional protections as the content of a newspaper editorial. What instruction we have received from the Supreme Court suggests that whatever expressive freedoms are incident to cable construction must be balanced against competing societal concerns. This is perfectly consistent with established First Amendment law. Those who market securities must conform what they say to SEC rules. Those who picket must honor the NLRB. Newspapers which merge commercial operations must have their editorial independence approved by the Justice Department. Government workers are constrained by the Hatch Act from partisan political activities. All of this conduct involves expressive freedoms, but all are balanced against competing societal interests. Who can better identify the relevant interests involved in a particular overbuild than a fully-informed local franchising authority? Suppose it finds that the overbuilder is attempting to "greenmail" the incumbent? As a St. Paul court ruled, such an applicant has no right to rattle the First Amendment sabre.[20] Suppose it finds that an overbuild will undermine universal service? Other courts have suggested that would sustain denial.[21] No doubt franchising authorities are political bodies; but a well conducted local airing of the facts of overbuilding would go a long way toward identifying and evaluating the public costs which most overbuilders would rather keep hidden.

1/ A classic example is the San Francisco federal district ruling in Century Federal, Inc. v. City of Palo Alto, No. C-85-2168 EFL (N.D. Cal. Sep. 1, 1987 and Oct. 12, 1988).

2/ See, e.g., Central Telecommunications, Inc. v. TCI Cablevision, Inc., 800 F.2d 711 (8th Cir. 1986), cert. denied, 107 S. Ct. 1358; Omega Satellite Products Co. v. City of Indianapolis, 694 F.2d 119, 126 (7th Cir. 1982); Nor-West Cable Communications v. City of St. Paul, No. 3-83 CIV 1228 (D. Minn. Sep. 1, 1988).

3/ Erie Telecommunications, Inc. v. City of Erie, 659 F. Supp. 580 (E.D. Pa. 1987), aff'd, 853 F.2d 1084 (3d Cir. 1988).

4/ In the Jefferson City case, the Supreme Court refused to review a ruling that the natural monopoly characteristics of cable television justified exclusive franchising. 107 S. Ct. 1358. In the Los Angeles case, the Supreme Court remanded an overbuild case for a "balancing" of competing interests. City of Los Angeles v. Preferred Communications, Inc., 106 S. Ct. 1034 (1986).

5/ Micro-Cable Communications Corp. v. Telesat Cablevision et al, No. 87-8358-CIV-SCOTT (S.D. Fla. filed 1987).

6/ Telesat Cablevision, Inc. v. City of Riviera Beach, No. 87-8207-CIV-MARCUS (S.D. Fla. filed Mar. 13, 1987) and GWC 104, Inc. v. City of Riviera Beach, No. 87-8208-CIV-MARCUS (S.D. Fla. 1987).

7/ See, e.g., Centel Cable TV v. Admiral's Cove Associates, 835 F.2d 1359 (11th Cir. 1988); NYT Cable TV v. Homestead at Mansfield et al, A-42/43/44-87 (N.J. Sup. Ct. June 28, 1988); Salvaty v. Falcon Cable Television, 165 Cal. App. 3d 798, 212 Ca. Rptr. 31 (2d Dist. 1985).

8/ See, e.g., Micro-Cable, supra n.5 (anti-trust); Wilco Electronic Systems, Inc. v. Gary Davis, No. 02307 Philadelphia, 1987 (Superior Ct. PA June 8, 1988) (1st Amendment); Cable Holdings of Georgia Inc. v. McNeil Real Estate Fund VI, Ltd, 678 F.Supp. 871 (N.D. Ga. 1986) (scope of easement).

9/ P. Dempsey, The Dark Side of Deregulation: Its Impact on Small Communities, Administrative Law Review 445 (1987); Deregulation Raises Prices, Cuts Services In Many Rural Areas, Wall Street Journal A-1 (Oct. 5, 1987).

10/ E.g., Touche Ross, Metro Dade Report on Overlapping Cable Franchise Study, Oct. 7, 1987; Booz Allen, Analysis of Ovderlapping Franchises, Oct. 9, 1979; Malarkey-Taylor, Economic Analysis of Cable System Overbuilds, Jan. 1987; Do Overbuilds Make Sense?, Cable Marketing 34 (Nov. 1987).

11/ For example, after overbuilds collapsed in Manatee County FL, basic rates jumped $6.00; in Ft. Oglethorpe GA, basic rates jumped $6.00; in Polk County FL, basic rates jumped $2.00.

12/ E.g., American TV and Communications v. Truvision, J87-0181 (L) (S.D. Miss. Nov. 29, 1987); Cablevision Industries Inc. v. Telesat Cablevision Inc., No. CI-87-1177 (Orange Co. Cir. Ct. Mar. 1987).

13/ E.g. Mobile Cablevision v. Group W Cable, Inc., Civ. No. CV 86-000145 (Mobile Co. Cir. Ct. April 21, 1986).

14/ P. Kagan Assoc., Cable Television Franchising, Nov. 30, 1987.

15/ In Lake Charles LA, the overbuilder declined to accept such a franchise in 1988. In Panama City Beach FL, the overbuilder balked at a $500,000 performance bond, negotiated a reduction to $100,000, and then sold out to the incumbent, all within three months of his initial committment.

16/ Touche Ross, supra n. 10.

17/ Clay Co. FL rejected an overbuild in 1988 after discovering that in 14 of the 18 communities cited in a Pasco County "study" of overbuild success, the cable overbuild resulted in only one surviving or in rates either increasing or failing to decrease.

18/ A. Smiley, Direct Competition Among Cable Systems, EAG 86-9 (U.S. Dept. of Justice, Antitrust Division, June 5, 1986).

19/ E.g., Fla. Stat. Ann. § 166.046 (1987); Tenn. Code Ann. tit. 7, ch. 59, § 1 (1988).

20/ Nor-West, supra n.2.

21/ Pacific West Cable Co. v. City of Sacremento, 672 F. Supp. 1322 (E.D. Calif. 1987).

Douglas Gomery
4817 Drummond Avenue
Chevy Chase, Maryland 20815
26 September 1988

The Economics of the "New" Hollywood

During the 1980s we have read over and over about a "new" Hollywood. Examples of change oft-cited changes include:

- New owners, the most famous being Rupert Murdoch and his takeover of Twentieth Century Fox.
- New production companies, such as New World and Cannon mounting serious challenges to the Paramounts and Columbias.
- New national exhibition chains. For example, Cineplex Odeon controlled less than one hundred screens in 1980 and today holds more than fifteen hundred.
- New stars, principally crossing over from television; possibly today's three hottest male movie stars initially made their names on the small screen: Michael J. Fox, Tom Hanks, and Eddie Murphy.
- New technologies, such as cable television, satellite to home delivery, home video, and high definition television -- all of which deliver more movies to more fans than ever before.

The American movie industry has survived all these changes, indeed prospered, even as pay cable and home video ancillary markets have grown into billion dollar industries.

I want to examine in this paper the fundamental economic principles underlying the American film industry. I shall argue that although there have been new owners, new entrants, new attractions, and new technologies, Hollywood as an economic institution remains essentially unchanged. It is organized and operates on the same principles familiar to industry watchers since the late 1920s.

THE OLIGOPOLY

The old Hollywood was labeled an evil trust by its opponents. This oligopoly consisted of the major companies, all of which owned theatre chains (Paramount, Loew's/MGM, Twentieth Century-Fox, Warner Bros., and RKO), and three smaller producer-distributors (Universal, United Artists, and Columbia.) Billionaire Howard Hughes dismantled RKO thirty years ago, but the other members of the so-called Big Five and Little Three are doing just fine. MGM and United Artists, at least as of now, are one. And Disney, long a minor player, has been revitalized into a major force, indeed the most successful company over the past eighteen months.[1]

The oligopolistic structure of the American film industry has been remarkably stable over the past sixty years, weathering the coming of sound, the coming of wide screen and color, the coming of television, the coming of cable television, the coming of home video, and a host of other technological changes which many have predicted would alter (or even dismantle) this basic structure. The seven major Hollywood studios may not be America's biggest companies, but they surely are among the most adaptable, agile and durable.

---

[1] For a survey of the Hollywood oligopoly of the 1930s and 1940s see Douglas Gomery, The Hollywood Studio System (New York: St. Martin's, 1986), pages 1 - 11.

Although seven companies dominate the industry, there have been changes in ownership. But I argue that these changes have had no effect on the basic structure of the industry, a structure which would look familiar to a Louis B. Mayer or any of the brothers Warner. Eight companies dominated in 1930, and in 1950. Today there are seven. Only RKO has been dissolved (as a producer), Disney has arisen to replace it, and MGM and United Artists have merged.

Today, old-fashioned entrepreneurs own and operate two of the seven "majors" -- Twentieth Century Fox and MGM/UA. These modern movie titans' reach vastly exceeds anything ever dreamed of by past movie moguls. Rupert Murdoch took control of Twentieth Century Fox in 1985, and then combined it with the chain of six big-city independent television stations that he acquired from Metromedia Television. Since then he has concentrated on building a fourth television network.

Kirk Kerkorian took charge of MGM in the late 1960s, and merged it with United Artists in the early 1980s. In a stunning turn of events in 1985, Kerkorian temporarily sold the MGM "half" to Ted Turner who desperately wanted the classic MGM movies to show on his SuperStation. Few were surprised when a year later Turner sold MGM back to Kerkorian. Another sell off nearly took place last summer. Kerkorian has proved the master at making millions selling and reacquiring his two movie companies.[2]

---

[2] For background on Murdoch's foray into the film business see Douglas Gomery, Vertical Integration, Horizontal Regulation -- The Growth of Rupert Murdoch's Media Empire," Screen, Volume

In contrast the other five majors represent billion dollar, multi-national media conglomerates, in some state of diversification. The most successful through the 1980s have been Paramount Pictures, a division of Gulf + Western, Warner Bros., part of Warner Communications, MCA's Universal, and the Disney studio. All generate considerable profits from a wide variety of mass media enterprises including theme parks (MCA and Disney), recorded music (MCA and Warner), publishing (Paramount and Warner), and television production (all four). In particular, Disney, once the new management team of Michael Eisner, Frank Wells, and Jeffrey Katzenberg took command in 1984, has aggressively claimed the status of a major studio. The most recent figures indicate that it should surpass Paramount and move into first place in the studio domestic box-office rankings for 1988.

The remaining major, Columbia Pictures, went through most of the 1980s as a division of Coca-Cola. Coke brought a research-oriented, marketing philosophy to a business where instinct and intuition had long proven the business methodology of choice. Coca-Cola tested a new philosophy for making money from movies. It tried to harness the creative energies in Hollywood by having traditional M.B.A.'s formulate strategies for optimizing sales and rentals from "movie software." Despite a series of attempts

---

XXVII, Number 4 (May - August 1986), pages 78 - 87. For a unique look into the usually secretive business dealings of Kirk Kerkorian see Aaron Latham, "Requiem for Leo the Lion," Manhattan, inc, Volume V, Number 9 (September, 1988), pages 98 - 106.

at cross-promotional advertisements for soft drinks and movies (consider the case of Diet Coke and Steve Martin's Roxanne), Coke never was able to create a blockbuster film, and in 1988 spun off the company into today's Columbia Entertainment.[3]

In Hollywood parlance these seven corporations -- Twentieth Century Fox, MGM/UA, Paramount, Warner, Universal, Disney and Columbia -- represent the latest incarnation of the Hollywood oligopoly. Every few years a couple of bold pretenders (recently, Cannon and New World Entertainment) emerge to challenge the majors, and capture the headlines of the business press on the basis of one or two modest hits (or even the promise of such). But no challenger has survived over the long run. In contemporary Hollywood, there are dozens of independent producers. But none are truly independent since all require distribution through one of the seven majors.

VERTICAL INTEGRATION

The United States Supreme Court found in the famous Paramount Case of 1948 that one of the pillars of power of the "Old Hollywood" was its vertical ownership and control of theatre chains throughout the United States. It forced the Big Five to sell its theatres. And for the years from the early 1950s through the early 1980s, Hollywood stayed out of the theatre business, at least in the United States. That changed several years ago when Paramount acquired the Trans Lux chain, and the deal cleared the Department of Justice. Other takeovers soon followed. Today Hollywood is again the significant player in the theatre business throughout the United States.[4]

---

[3] For a good background on the operations of the majors see Peter J. Dekom, "The Motion Picture in 1987: A Study in Economic Turmoil," American Premiere, Volume VIII, Number 1 (Spring, 1988), pages 9 - 15. Updates were gathered from the pages of weekly Variety.

Paramount and Warner Bros. (assuming the deal goes through) own the Mann chain (the strongest in California and indeed the far west), the Famous Players Canadian chain, and some smaller circuits in the East. At the beginning of 1988 Mann itself was the sixth ranking chain in size in North America, and Famous Players Canadian was eighth.

Columbia Entertainment owns the rapidly growing Loew's chain. Loew's controls some five hundred screens and is looking toward one thousand by the early 1990s.[5]

But the deal which made the greatest headlines occurred when Universal took half interest in Cineplex Odeon. Indeed the agreement between Garth Drabinsky, founder and driving force of Cineplex Odeon, and Lew Wasserman, long considered the most powerful boss in Hollywood, rocked the industry. The

---

[4] See Gomery, The Hollywood Studio System, pages 10 - 12 plus the chapters on the individual corporations for a description and analysis of the vertical integration in the old Hollyw

[5] A basic survey data for theatre ownership can be found in Variety. This information is from the 20 January 1988 issue, pages 37, 68. More mergers have taken place in 1988, and continue to the present. Rumors swirl about the sale of the third largest chain, American Multi-Cinema, and the fourth largest General Cinema. Interestingly the largest, for the moment, is United Artists Theatre Circuit, owned by TCI, the nation's biggest cable television franchiser.

Universal/Cineplex chain owns and operates more than sixteen hundred screens, and brags that it is the largest in North America.

It is certainly the most aggressive. Drabinsky has long boasted that he wants Cineplex to dominate the top movie-going markets in the United States. (Cineplex already controls Canada.) To accomplish his goal, during the past three years Drabinsky has purchased theatres in Los Angeles, Washington, D.C., Chicago, Seattle, and New York City. In less than a decade Cineplex Odeon has grown from a Canadian art house circuit to the most powerful chain in North America. To expand so rapidly Drabinsky raised the necessary millions by selling shares first to the Brofman family in Canada and then "half" to MCA. This latter deal provided Universal with direct access to some of the best situated multiplexes in North America, creating a system of vertical integration not seen since 1950.[6]

Drabinsky has spent millions making sure movie fans realize the difference between going out to the movies in one of Cineplex's theatres, and staying home and watching television. The company differentiates its product by offering a return to at least some of the elegance of movie-going during days of the picture palace. Advertisements stress wide, comfortable seats,

---

[6] For more on Cineplex-Odeon see Variety, 25 November 1987, pages 7, 32; Variety, 2 December 1987, pages 5, 15; Mara M. Balsbaugh, "Cineplex Odeon Corporation," (New York: Smith Barney Research, 1987), pages 3 - 33; The Wall Street Journal, 7 May 1987, page 59. See also Lisa Gubernick, "Movie Madness," Forbes, 8 February 1988, pages 37 - 38.

"real" butter on the popcorn, and standards for sound and image seemingly long forgotten.[7]

What Hollywood has come to realize in the 1980s is that the "theatrical window" is still where blockbuster hits are developed. Feature films begin their marketing life in theatres, and to maximize returns in the ancillary markets of pay-cable, home video, foreign distribution, and over-the-air television, a film must do well at the theatrical box-office. The multiplex is where the twenty-million dollar investment (for the average theatrical feature) is made or broken. A theatrical blockbuster guarantees millions of additional dollars from the home video, and pay television arenas. That is why the major Hollywood companies work so hard to craft a hit in the theatres. Once they have a proven commodity there, the rest of the way is smooth sailing.

But the continuing importance of theatres has only become obvious in the last few years. Consider that as we turned into the 1980s a number of industry analysts studied the film industry, and predicted that there probably would no need for movie theatres by 1990. Everybody could (and would) stay home, and view Hollywood's best through the new television technologies. Instead what has happened is that going out to the movies has remained viable (increasing slightly) even as pay-cable and home video grew into billion dollar industries.

---

[7] Garth H. Drabinsky, "New Strategies for the Future: Two Challenges Facing Exhibition," Boxoffice, Part 1 in February, 1988, pages 10 - 11, and Part 2 in March, 1988, pages 81 - 82.

During the 1980s more new movie theatres have been built, creating more theatrical screens in the United States than anytime in film history. Theatrical release requires more and more theatre screens so Hollywood can take full advantage of the economies of television advertising. Even though the cost of marketing a film can often exceed ten million dollars, as it is spread over more theatres, marketing costs per theatre per film remain relatively low. .

The economies of scale provides the foundation of the multiplex. A complex with twelve screens is more flexible to be able to maximize revenues from a hit through careful placement of hit films and wide spread television advertising. The blockbuster can be shifted to the largest auditorium, a less popular film to a smaller one. Economies of scale of operation also allow the owner of a theatre (now a Hollywood company) to garner additional profits from concession sales. Popcorn sales alone in the 1980s consistently rang up more than three hundred million dollars per year. A well-run concession stand, it has been found, can generate as much as seventy-five cents of profit for each and every ticket buyer.[8]

PRICE DISCRIMINATION

In the old days the Hollywood majors through their vertical ownership set the prices, and through careful use of runs, zones and clearances, were able to maximize revenues from box-office take. In other words they worked together to enforce a system of classic price discrimination. A generation ago the process was simple. Features opened with big premieres and publicity. Then they played off from first to second to third-run. The coming of television simply added two runs at the end. A film earned the bulk of its money from play in theatres; television, even with a sale for showing on one of the three major television networks, accounted for less than a tenth of total revenues.[9]

---

[8] The trade paper, Boxoffice, in its massive March, 1998 special edition provides a candid survey of the operations of the movie theatre business, and testimony to the effects of vertical integration.

This classical pattern of runs was significantly expanded in 1975 with Time, Inc.'s innovation of a pay-cable service, Home Box Office. In the 1980s the home video revolution offered a further set of runs (day-and-date with pay-cable usually). Indeed the Hollywood majors have been the big winners in the world of the new television technologies because they have skillfully learned to take advantage of new possible windows of release to extract more revenue from their features films through a careful use of price discrimination. They have adjusted their behavior quickly and efficiently, and stand today with more power and profits than ever before.[10]

---

[9] For an analysis of the price discrimination in the movie industry during the 1930s and 1940s see Gomery, The Hollywood Studio System, pages 14 - 22. See also Louis Philips, The Economics of Price Discrimination (New York: Cambridge University Press, 1983).

[10] There are further changes on the horizon. Pay television operators want to collect more money so they are implementing a system to "premiere" blockbuster films on a one time basis before

Today feature films still begin their marketing life in theatres. As noted above, the first-run theatre remains the principle "voting booth." If one can fashion a theatrical blockbuster during the crucial first weekend of release, then one should reap added revenues from the home video, and pay television arenas. Even as recently as ten years ago, multiplexes supplied more than three-quarters of the revenue for an average Hollywood feature film. Today theatres provide less than half. Movie theatres are doing well, but home video is doing far better.

The financial data from the home video market for movies is indeed impressive, even extraordinary. In 1980 the Hollywood majors collected some twenty million dollars from world wide sales of video cassettes. In 1987 the figure was over seven billion dollars. Ten years ago an average film took in nothing extra from home video; today that revenue contribution usually amounts to one-third.

In the mid-1980s, outlets renting video tapes seemed to be popping up everywhere. Grocery stores, drug stores, convenience chains, gas stations, dry cleaners and even in some locales, pizza delivery services, were willing to lease tapes. Consider that in 1987 fully a sixth of all tapes were rented from grocery stores and drug stores alone. Hollywood has sought ways to participate in this rental market by establishing different

---

the film reaches the video store. This Pay-Per-View system represents another window to extract dollars. The constraint seems to be the ease of ordering and billing.

prices of the tapes through time. Initially a tape is sold for close to one hundred dollars, with principle buyers being video rental stores. But once that market has been exhausted, then a special sale is held for about twenty dollars, aimed at individual buyers.[11]

But home video hardly stands alone in Hollywood's ancillary markets. In 1982 HBO seemed to be on the verge of establishing itself as an eighth Hollywood major. But in recent years pay cable revenues in general, and HBO in particular have begun to level off. Too many folks seem to prefer to rent the cassette when they want to see it, not wait until HBO schedules the film. Still HBO continues to dominate the multi-million dollar pay-cable market.[12]

With repeated showings on pay-cable and home video, theatrical features shown on network television, has fallen to lows unmatched since Twentieth Century Fox and NBC introduced "Saturday Night at the Movies" twenty-five years ago. Now

---

[11] The latest innovation to assist the majors in collecting more money is called Pay-Per-Transaction or PPT. As currently being explored, PPT would allow a store to lease its movies from the studios rather than buy them outright. Instead of paying a one time fee, the video store would split (usually fifty-fifty) with the studio the income each time the tape is rented. PPT allows the video renting agent to carry more tapes, and thus satisfy often frustrated customers, and keep them coming back. For a compact description of various marketing innovations see Jennifer Stern, "The Case of the Missing Movies," Video Review, October, 1988, pages 23 - 26.

[12] Janet Stilson, "HBO: From Polkas to Prize Fights," Electronic Media, 30 November 1987, page 22 - 24; Chuck Ross, "Inside HBO's Film Scheme," Television Radio/Age, 26 October 1987, page 47 - 49.

"Saturday Night at the Movies" (or a counterpart on another night) means a series of ubiquitous TV Movies. If ever there was proof that the TV movie had replaced this particular run for theatrical features, it came on a Sunday night in February 1984 when CBS first aired Star Wars. CBS was so confident of high ratings that it doubled its prime-time advertising rates. To nearly everyone's surprise Star Wars was beaten in the ratings by Lace, a steamy made-for-TV movie which cost only three million dollars to make, less than half what CBS paid to rent Star Wars.[13]

The final run for a movie is its presentation by an individual television station, most frequently a non-network station, delivered either over-the-air or through cable. Today such independent stations represent an increasingly important market. It usually takes a couple of years for important feature films to make their way to independent television (less for box-office failures), but once released, the film is shown over and over again.

Turner Broadcasting Service's SuperStation is one of the most watched of all cable selections in the United States. Fully a third of WTBS's time is filled with old films. In 1985 Turner purchased MGM, not for its current productions, but for access to its library of old films. Turner so wanted the old films he paid far more than he could afford, and thus had to sell the remainder of MGM back to Kirk Kerkorian a year later. And Turner's SuperStation is no different from over-the-air independent stations found in all sizable U. S. cities. Into every television home, a vast array of Hollywood sound films play, day after day. Indeed to make the product "different," Turner has begun to colorize older black & white features. The transformation is not cheap, sometimes costing up to a quarter of a million dollars. Turner has led the way with his presentations of They Were Expendable, Action Across the North Atlantic, The Maltese Falcon, and Yankee Doodle Dandy.[14]

---

[13] Douglas Gomery, "Hollywood's Business," The Wilson Quarterly, Volume X, Number 3 (Summer, 1986), pages 53 - 56; Variety, 15 July 1987, pages 37, 54; The Wall Street Journal, 21 February 1984, page 30.

Whatever additional new television technologies appear in the future, the business of Hollywood will continue to be to seek additional ways to maximize revenues from any and all films. It will do this by trying to formulate policies for effective price discrimination by segmenting potential customers and having them pay as much as possible for a viewing, starting with a theatrical showing, then home video, pay-cable, network television, and local television. The latest wrinkle in this formula is Pay-Per-View, which would come after a theatrical release and before home video and pay cable. So far PPV has not produced great results, for movies, but the process of innovation continues.[15]

---

[14] Variety, 15 July 1987, pages 5, 27; Television Radio/ Age, 28 September 1987, pages 26, 30; Variety, 15 April 1987, pages 43, 206; Susan Linfield, "The Color of Money," American Film, January/February 1987, pages 29 - 32.

WORLD-WIDE DISTRIBUTION

In the old days the majors held dominion for distribution throughout the world. Everywhere one went, Hollywood films dominated . That has not changed. Today Hollywood still holds hegemony through most of the world. And its trade association, the Motion Picture Association of America (MPAA) spends a great deal of time making sure that economic network is maintained. Indeed, in today's shrinking world trade for the United States, the movie business is one of the few positive signs on the United States trade balance sheet.

Year in, year out the majors control almost eighty percent of the movie business in the United States and approximately half the market in Sweden, West Germany, and several other nations in Western Europe, not to mention Asia. Hollywood derives roughly one-third to two-fifth's of its theatrical box-office revenues from overseas. This has been the long-term "secret" to the majors' success.[16]

---

[15] The trade press is constantly filled with reports of the success of failures of Pay-Per-View. See, for example, Electronic Media for 23 May 1988, or 4 July 1988, or Multichannel News for 27 June 1988 or 4 April 1988.

[16] Variety is constantly reporting of the success of the American film around the world. In 31 August 1988 issue it reported (pages 5, 34) that "the foreign releasing arm of Paramount, Universal, and MGM/UA, equalled its best year ever in the 12 months ending June 30 [1988]." These three companies operate jointly around the world to keep cost to a minimum.

THE FUTURE

If the present power of the Hollywood majors stems from a simple transformation of techniques learned sixty years ago, is there any factor on the horizon which might significantly change these basic patterns of structure and conduct? Actually open it up to new competitors? Or rid the oligopoly of some of its power? Or defuse the vertical integration? Or have prices set competitively? Let me suggest two possibilities.

1. For years the coming of some new technology was predicted to fundamentally finally alter Hollywood's hold on the marketplace. It seemed only to be a simple matter of economic substitutes. With pay-cable, home video or today's Pay-Per-View, who would want to go out to the movies? Millions, we were told, would stay home, and abandon the theatre. Many do stay home, but this summer millions paid a record at the box-office to go out to the movies.[17]

The latest candidate which some claim will strike the death knell of the movie theatre is high definition television with its superior, wide screen video image. I doubt this. First, the film image is also improving in quality, and will remain vastly superior to even the high definition television image. And second, image quality does not seem to matter much. If it did, why did the technically superior Beta format for home video fail so badly in market competition with VHS?

---

[17] See Variety, 14 September 1988, pages 3, 45.

2. Exogenous shocks could also have an influence. I have argued elsewhere that television did not cause the decline in theatrical movie attendance from 1945 onward. Rather the transformation came about because of an extraordinary shift in the location of where the middle class movie audience lived (the move to the suburbs), and the size of families (the baby boom). Hollywood eventually adjusted, and became the principle supplier of programming for television.[18]

What outside shocks might be in the cards? A major recession in the United States could cripple the theatre business and hurt Hollywood through its sizable investment in theatres. Or the movie theatre might be abandoned by the members of the baby boom generation, long the movies most faithful of fans. They might turn exclusively to television, or something else. Such shifts would require a reshuffling in Hollywood.

Because Hollywood, however, is so well diversified, I do not see that either of these exogenous shocks doing long run damage. My best guess is that Hollywood will be structured and operating as we have long been accustomed to it well into the next century, utilizing the principle described and analyzed in this paper.

---

[18] Douglas Gomery, "The Coming of Television and the 'Lost' Motion Picture Audience, Journal of Film and Video, XXXVII, Number 3 (Summer, 1985), pages 5 - 11.

THE OPPORTUNITY COSTS OF SPECTRUM ALLOCATED TO HIGH DEFINITION TELEVISION

Dale N. Hatfield
Gene G. Ax

The Sixteenth Annual
Telecommunications Policy Research Conference

Airlie House
Airlie, Virginia

October 30-November 1, 1988

Hatfield Associates, Inc.
4840 Riverbend Road
Boulder, Colorado 80301
(303) 442-5395

I am pleased to have the opportunity to participate once again in this Airlie House Telecommunications Policy Research Conference. I am particularly pleased to participate in this session on High Definition Television (HDTV). However, I am more than a little apprehensive about the topic I have been given; namely, the opportunity cost of accommodating an Advanced Television Service in the existing VHF and UHF television bands.

I am apprehensive because I am currently serving on a pro bono basis as the Chairman of the Spectrum Utilization and Alternatives Working Party of the Planning Subcommittee of the FCC's Advisory Committee on Advanced Television Service (ATS). In that capacity, I have tried to maintain a degree of impartiality and objectivity and I have tried not to become an advocate of any particular use of the spectrum that might be available. Here, however, I have been cast in the role of talking about alternative uses of VHF and UHF television spectrum that might otherwise be used to accommodate a new Advanced Television Service. I intend to pursue that role vigorously.

This all means that, while I will necessarily be talking about the economic and other benefits of allocating the available spectrum for purposes other than ATS or HDTV, I am not advocating such allocations or assignments in lieu of its use for the latter purpose. In short, I remain "officially" neutral.

Before I turn to the main part of my talk -- which I intend to keep very brief -- perhaps I should describe what is meant by the term opportunity cost. Opportunity costs arise in any situation where you are faced with economic choices. It is the cost of using limited resources for one thing rather than another. If I go out and pay cash for a new car, I am foregoing the interest income that I could have earned by keeping the money in the bank. By attending this conference today and tomorrow, I am foregoing a certain amount of income from my consulting practice. In either case, the additional income that I am foregoing is the opportunity cost of taking the alternative action. In the case of buying a new car, the interest income I forego is part of the cost of owning the car, and it is just as real as buying the gas and oil to run it.

Spectrum is a scarce resource and if we use it for one thing, we deny its use for another. Thus there are opportunity costs associated with the allocation of spectrum for an over-the-air HDTV system. Ideally, it would be nice to quantify the value of these alternative uses of the spectrum in order to estimate the opportunity cost of the allocation to HDTV. This would provide valuable information with which to address the threshold issue of whether the spectrum should be allocated for HDTV. However, I can't even begin to identify all of the alternative uses, let alone quantify the economic value of each. Nevertheless, what I will try to do in the paragraphs which follow is identify what I think are possibly the highest value other uses, and I will quantify the value in the few cases where I have some hard information.

Additional Broadcast Outlets

Perhaps the most obvious alternative use of the spectrum would be to create additional conventional (i.e., NTSC) television assignments. For example, its recent Tentative Decision and Further Notice of Inquiry in the Advanced Television Systems Proceeding, the Commission explicitly stated that it was agreeable to allowing a licensee to transmit a normal NTSC signal on his/her regular channel and, simultaneously, an incompatible HDTV signal on a second 6 MHz channel. This suggests the possibility of transmitting a second signal that is independently programmed by a different licensee. More generally, of course, in a given region, there is a tradeoff between providing higher quality signals from existing stations and adding more stations with the same quality signals.

In its decision, the Commission notes this possibility by stating "Assuming the use of more bandwidth permits transmission of better quality signals, we must compare the costs and benefits of using spectrum for enhancing television performance with increasing the number of NTSC stations..." Besides the obvious economic value of additional stations in a market like Los Angeles, there would be additional public policy benefits associated with additional programming diversity and additional broadcast station ownership opportunities for minorities and other under-represented groups. Unfortunately, I do not have the information necessary to estimate the economic value of additional television stations in the larger markets, so I can do no more than mention it as an opportunity cost of allocating spectrum for HDTV.

Cellular Mobile Radio

The growth of cellular mobile radio systems has greatly exceeded the expectations of many observers. Because of this growth, the cellular carriers have put pressure on the FCC to allocate additional spectrum to their service. The Commission responded once to this pressure and allocated an additional block of 10 MHz to the original 40 MHz allocation. The total of 50 MHz of spectrum is divided between two carriers in each market. The spectrum lies in the 800/900 MHz region.

The cellular mobile radio service gets its name from the underlying concept which involves dividing a geographic region (e.g., a metropolitan area) into a series of smaller geographic areas called cells. Base station transmitting and receiving equipment is placed in each cell and connected by wirelines or microwave to a central switching computer and from there into the regular telephone network. Each cell is equipped with multiple channels and the associated base station equipment provides coverage over the cell. These cells are small enough, by virtue of the low power and low antennas used, to allow the same radio channels to be used simultaneously in several cells in the region.

One key to the cellular concept is that, in an initial system covering a metropolitan area, the cells can be quite large, and as demand develops, these large cells can be subdivided into smaller cells. When the cells are large, a particular channel may only be used once or twice in the region; in the later stages of development, the same channel may be used ten or more times.

Thus one can increase the capacity of a cellular system in two ways: (1) by adding smaller cells and simultaneously reducing the size of existing cells or (2) by simply adding spectrum (channels) to the existing sites. For example, to double the capacity of an existing system, I could double the number of cells (by going to more and smaller cells - a process called cell division) and thus double the number of times each individual frequency is used in the region. Or, in the alternative, I could double the capacity simply by doubling the amount of spectrum I am using.[1] Adding spectrum delays the point at which added cells are needed.

If spectrum were reallocated to the cellular service instead of to HDTV, what would its value be? Or stated another way, what would a cellular carrier be willing to pay to avoid "cell division?" Fortunately, we have enough information to be able to compute the amount of additional investment in new cell sites that can be avoided by adding an incremental amount of spectrum, say 1 MHz. A cellular carrier should be willing to pay this amount to avoid or postpone the investment in additional sites. This amount can be used to estimate the opportunity cost of allocating an additional 1 MHz of spectrum for HDTV rather than to the cellular radio service.

---

[1] I am ignoring trunking efficiency and certain other effects in this example, but the concept is correct. Note also that there are practical limitations on both how large and how small the cells can be. The carriers argue that certain practical limitations keep them from making cells with a radius of less than one mile or so.

Gene Ax of my consulting firm has actually carried out the necessary calculations and the results are shown in the accompanying table. Details of the methodology used and the supporting calculations are given in the appendix. Before I make some comments on the results shown in the table, it might be useful to make a few comments on the economics of cellular systems. First, there are not strong economies of scale in a cellular system once you get beyond the start-up stage and frequency reuse becomes necessary. Second, when you add capacity by adding channels (i.e., spectrum) at each cell site, you are adding just variable costs -- the costs of additional transmitters/receivers, telephone lines, etc. You avoid all of the one-time costs of acquiring the site, building the equipment shelter, constructing the antenna tower, etc., and you gain from the additional efficiencies associated with having more channels at a given site.

In contrast, when you add capacity by adding cell sites, you not only absorb these same variable costs, but you incur the additional fixed cell site costs and you do not gain the efficiencies associated with additional channels at an existing site. This means that it is less costly to gain capacity by adding spectrum, if it is priced at zero. What we are estimating here is the price that the cellular carrier should be willing to pay if spectrum were "for sale."

Note from the table that we estimate that the value of an additional one megahertz of spectrum to a cellular operator needing additional capacity ranges from about $300 K for a small seven cell system to nearly $7 M for a 224 cell system. The value of an additional one megahertz increases with the number of cells in the system because of the larger amount of frequency reuse. For example, if the spectrum is reused 10 times in a region with many small cells, an additional 1 MHz of spectrum actually adds 10 MHz worth of capacity.

Gene included the values for 100 cell and 170 cell systems in the table because the former is the number of cells that PacTel had in Los Angeles in August, 1988, and the latter is the number of cells they are projected to have by the end of 1989. The corresponding value of one megahertz of spectrum in these two cases is about $3 M and $5 M respectively. Thus a single, 6 MHz television channel that might be utilized for HDTV in Los Angeles will have an opportunity cost of roughly $30 M in 1989. If nothing else, our calculations indicate the high financial stakes that are involved with the allocation of spectrum for an Advanced Television System.

Trunked Mobile Radio

At the same time the FCC originally reallocated spectrum from UHF television to create the Cellular Mobile Radio Service, it also reallocated spectrum for private land mobile radio systems. Private land mobile radio systems have historically been used for private dispatch communications rather than mobile telephone service per se. Dispatch type communications are used by police departments, businesses (e.g., a plumbing and heating shop), and other entities for communications between a dispatcher and a fleet of vehicles and personnel in the field.

In reallocating spectrum for private land mobile radio uses, the Commission wanted to encourage the development of a particularly efficient technology called trunking. In modern trunked land mobile radio systems, access to a large number of radio channels (say a pool of five to 20) is controlled by computer logic that gives the user a channel from the pool for the duration of his or her message, and then returns it to the pool for use by other users. In other words, efficient sharing of a number of channels is promoted because particular users or groups of users that do not have heavy demands for communications during a particular period or moment in time can, in effect, loan their spectrum capacity to others who may be experiencing a peak in demand and vice versa. Trunked radio systems are much more efficient in their use of spectrum than manual systems involving users being limited to a single channel. In order to be able to reach a number of field units simultaneously, trunked radio systems use relatively high power and high antennas to cover an entire metropolitan area from a single site.

In order to encourage the use of this advanced technology for the provision of dispatch communications, the Commission created a new service known as the Specialized Mobile Radio (SMR) Service. Licensees (i.e., operators of trunked radio systems) in this service provide land mobile radio communications services on a commercial, for-profit basis to entities eligible for licenses under the private radio rules of the Commission. Although it has received considerably less publicity in the popular and financial press, SMR operators have been extremely successful and the growth of the service has been almost as phenomenal as cellular. This growth -- and the growth of conventional private systems-- has also created pressure for additional allocations. At the same time it reallocated an additional 10 MHz of spectrum for cellular use, the Commission also allocated an additional 5 MHz of spectrum for use by SMR operators. It divided this reallocated spectrum into blocks of 10 channels each, where each channel consists of a 12.5 kHz-wide transmit and a 12.5 kHz-wide receive frequency. Thus there were a total of 20 blocks amounting to .250 MHz of spectrum each.

The Commission assigned these blocks using a lottery and the lottery winners are permitted to transfer (sell) their license after they have constructed the system. Through the activities of my consulting firm and with the help of an associate, Beverly Mang of Denver, we are aware of the recent selling prices of some of these systems. Since the systems are sold before they are loaded with users, it is possible to establish what people are paying --in effect -- for the underlying .250 MHz block of spectrum. In the largest metropolitan areas -- New York and Los

Angeles -- the price seems to be in the $500 K to $600 K range, or perhaps a little more. This translates into a current value per megahertz of more than $2 M. Thus the opportunity cost of using UHF spectrum in Los Angeles for one standalone channel of HDTV rather than trunked private radio/SMR systems is about $12 M based on current prices as we understand them.[2]

It is interesting to compare these marketplace prices with the corresponding opportunity cost that Gene computed for cellular. The last column in the table displays what the value of one of these spectrum blocks would be to a cellular carrier based on his analysis. Note that for a 100 cell system, the size of the PacTel system in Los Angeles in August, 1988, the imputed value of one block of trunked SMR 900 MHz spectrum was about $750 K. Thus the two values are surprisingly close. What entrepreneurs are paying for spectrum in Los Angeles today is roughly equal to the value of that spectrum to the cellular carrier. Lest I get into more trouble than I otherwise will from what I have already said, FCC rules currently prohibit the cellular carrier from buying this spectrum and converting it for use on their cellular system; and there may well be a host of good public policy reasons why they should not be allowed to do so as well.

---

[2] Note that we are assuming that the geographic coverage of the SMR system is the same as the coverage of the cellular radio system in Los Angeles. Because of the mountaintop sites typically used in the Los Angeles area for SMR systems, this is not an unreasonable assumption.

"Wireless Local Loop" Systems

In terms of long range public policy benefits -- if not direct economic value -- I think one of the best alternative uses of the available spectrum would be to allow the development of competition for telephone company local loops. Dr. Crandall of Brookings suggested this possibility at an early meeting of our ATS Planning Subcommittee. Some time ago, my consulting firm did some studies of the feasibility of using conventional cellular mobile radio systems to compete with ordinary local telephone services. A principal constraint to such competition, of course, is the availability of spectrum, especially in view of the somewhat unexpectedly large growth of cellular in the largest urban areas. Also, regular cellular systems are an overkill if all you want to do is provide "wireless loops" to portable or fixed telephones. It is overkill in an engineering sense because there is no need to build in large amounts of power margin for fading or the capability for handing off calls from cell-to-cell if the portable unit is only carried about on one's person.

As it turns out, the Europeans have already developed a portable telephone network that is called the CT 2 system. It can best be characterized as a high quality, public, cordless telephone system. It uses multiple frequencies, dynamic channel assignments, low power base stations and extensive frequency reuse to gain capacity. It is my understanding that the United Kingdom has already allocated 4 MHz of spectrum in the 900 MHz range for this purpose. This four megahertz of spectrum between 864-868 MHz provides for 40 channels, each 100 kHz wide.

The extremely high spectral efficiency offered by the system reportedly allows even the simplest implementation to serve user densities of over 5,000 users per square kilometer.[3] (This is many times greater -- perhaps two orders of magnitude -- than conventional cellular systems.) This is obviously adequate to provide substantial competition to the local telephone company and it has been confirmed by our own calculations. Indeed, by allocating several television channels in an urban area to this purpose, one could probably accommodate several competitors! Actual implementation of this type of system as a competitor to the local exchange carrier would require some method of getting from the low power base station site to a switching office. This could be done using short-haul microwave or existing feeder or other facilities.

In my opinion, the public policy benefits of such competition -- if it is feasible -- would be enormous. The need to regulate the Bell Operating Companies could be diminished significantly because of the presence of competition, and their ability to engage in cross subsidization, discrimination and other anticompetitive activities would be reduced. Thus the general public would not only benefit from a rather interesting new service, they could also benefit by a reduction in the direct and indirect costs of regulation. What is lacking is the spectrum; the technology is available almost "off-the-shelf."

---

[3] Clare Lees, "The Cordless Connection," Telecommunications, January, 1988, p. 63.

Other Potential Uses

The radio spectrum is an incredibly flexible transmission medium and its value has been greatly enhanced by the rapidly decreasing size and lower power consumption of solid state equipment and by the greatly increased functionality permitted by low cost computer logic. As a consultant, it is clear to me that many good spectrum consuming ideas never get very far because of the high cost of gaining access to the spectrum for a new service or application. In this section of the paper, I will mention some of these other uses, but it should be recognized that I am only scratching the surface. Indeed, the alternatives I will list here should be regarded more as indicative of the possibilities rather than an exhaustive treatment.

One candidate is an aeronautical mobile telephone service. The FCC is still struggling with finding a permanent home for the current commercial aeronautical mobile telephone service provided by GTE Airfone. Additional VHF/UHF spectrum would make it much easier and less controversial to accommodate this service and, perhaps, make it possible to have competitive suppliers. Another possibility is the land mobile satellite service. Several years ago, a number of applicants sought the opportunity to provide land mobile services via satellite using frequencies in the 800/900 MHz region of the spectrum. The Commission has forced these competing applicants to consolidate their efforts in a single system in a region of the spectrum higher than the one they originally sought. There were several advantages associated with the use of the spectrum they originally sought, including the possibility of subscribers being able to use a common radio and switch from the cellular mobile radio system to the satellite system whenever they were outside the coverage area of a terrestrial system. Such an arrangement would provide true nationwide mobile radio coverage. This opportunity, and the opportunity to accommodate competitive suppliers, is probably gone, but it illustrates some of the potential value of the HDTV spectrum.

Several years ago, the FCC denied advocates the spectrum necessary to establish a *personal radio communications service*. The VHF/UHF range is the ideal place to accommodate such a service because it is high enough to avoid long distance "skip" interference, but low enough to provide good coverage at moderate power levels with reasonably sized antennas. As originally envisioned, the service would have provided a much better quality of service than today's Citizen's Band service, and some observers liken it to a "poor man's cellular" since it would also have provided a rudimentary mobile/portable telephone capability. Based on the rapid and widespread acceptance of cellular, it is quite possible that this would have been a successful use of the spectrum. Nevertheless the Commission was unable to accommodate it in the 800/900 MHz band and there was never a real market test of the system.

In a recent paper delivered at a Columbia University sponsored conference on spectrum management, Henry Goldberg, a Washington, DC based attorney who practices before the FCC, related some of his experiences in trying to gain access to spectrum for clients. I was particularly struck by two examples that he gave of the difficulties of gaining access to spectrum for applications that would appear, at least on the surface, to have significant marketplace appeal. One example involved the Hewlett-Packard Company (HP). According to Mr. Goldberg:

> [HP] had been developing and experimenting with...a new wireless data terminal. The terminals, which were intended to be installed on movable robots, fixed computers, and automated assembly and transportation equipment, were designed to be used inside factories and warehouses for data collection and related purposes.

Off hand, this would appear to be a use of the radio spectrum that might produce significant productivity gains in factory and office alike but, as reported by Mr. Goldberg, the company "has failed to pursue further development of its wireless data terminals..." because the FCC has so far been unable to find a place in the spectrum to put them. It is my recollection that Motorola was also recently turned down in an attempt to gain spectrum for a wireless local area network with some of the same applications and advantages.

In the same paper, Mr. Goldberg also reported on the attempts by a client to gain access to spectrum for "an interactive television viewer response device that works with any kind of television transmission medium, over-the-air or cable."

VHF/UHF spectrum, of course, would be ideal for any of these devices, and I raise them only as current examples of applications that are being held back by lack of a spectrum allocation. It makes one wonder how many innovative, productivity enhancing, commercially successful services would be created if spectrum were readily available.

Summary and Conclusions

In this paper, the opportunity costs associated with the allocation of spectrum for High Definition Television have been explored and these costs have been quantified in a preliminary way for the cellular radio service and, based upon some marketplace prices, for the trunked, private Specialized Mobile Radio Service. The admittedly rough calculations for cellular and the marketplace confirmation from the sale of SMR systems indicate the potentially high opportunity cost associated with the deployment of over-the-air HDTV.

TABLE

INCREMENTAL VALUE OF ONE MEGAHERTZ OF SPECTRUM

No. of Cells in Cellular System	Incremental Value of One Megahertz of Spectrum	Imputed Value of Spectrum Used 10-Chl. 900 MHz Trunked Systems
7	$ 0.2 M	$ 53.5 K
14	0.4 M	$ 107.0 K
28	0.9 M	214.0 K
56	1.7 M	428.0 K
100	3.0 M	764.3 K
112	3.4 M	856.0 K
170	5.2 M	1,299.3 K
224	6.8 M	1,712.0 K

Appendix A

Calculation of the value of adding one MHz of spectrum to a cellular mobile system that is now limited to a total of 25 MHz

As indicated in the body of this paper, the purpose of this appendix is to provide enough information so that the reader can reconstruct the results of the Table concerned with the incremental value of one megahertz of spectrum to a cellular mobile system. A further purpose of this appendix is to describe the basic methodology that can be applied with new data and/or assumptions to come up with new results.

Assumptions

(1) The technology is fixed with the present 30 kHz analog FM channel systems having a total of 395 voice channels available within the 25 MHz of spectrum assigned to each of the two cellular systems in each market. (The newly developing digital cellular systems will require new calculations with new data and assumptions. These new systems will give greatly added capacity and more cost effective results over time, and, therefore will become attractive first in the major markets facing capacity constraints. Later on, these systems will become the system of choice in all markets as their improved capabilities and lower costs point the direction.)

(2) When cellular systems are in their early start-up phase in major markets, or in a mature phase in smaller markets, getting adequate coverage area for the customers is the main concern while having enough traffic capacity is only a secondary concern. Under these conditions it is cost effective to employ simple omni-directional cells (sites), cells that have fewer available channels when frequency reuse is required than the ones employing sectorized antenna patterns. However, if, and when, capacity becomes a concern as all the available channels are used up, and frequency reuse becomes a major requirement, the sectorized cells become more and more attractive. This is because more and more channels can be equipped at each site so that greatly increased capacity per site is achievable. Having greater capacity per site means that fewer sites are now necessary for supporting a given capacity, and this saves money by eliminating the large fixed costs that are associated with each site. Thus, there is always a "tug" between coverage and capacity requirements and associated costs so that most cellular systems in the major, high-capacity requirement markets employ a mixture of cell sizes and types of cells--sectorized or not.

Even though cellular systems employ a combination of 12, 7, and 4 cells (sites)--among other less likely possibilities--per each use of the set of available 395 "voice" channels, it is assumed that, for the high-capacity, mature, cellular systems employing significant reuse, the 7 cell, 3 sectors/cell combination represents an "average" or "typical" state for modeling the system as a whole. (As a rough "rule of thumb" one can estimate that each of the 395 available "voice" channels can support 25 subscribers. Thus, at least some reuse of spectrum (channels) becomes necessary in a cellular system when the number of subscribers exceeds about 10,000. For example, we can estimate that the Los Angeles PacTel cellular system, that is now supporting an estimated 170k subscribers, must be using each channel around 17 times overall, on average.)

(3) It is assumed that the fixed cost for every site (cell) is the same throughout the cellular system and that somewhat "lumpy" costs associated with the site, such as equipment mounting racks, buildings, heating, ventilation, and air conditioning, antenna systems, etc., can be adequately folded into "per-equipped-radio-channel" costs or investments.

(4) Throughout the range of cellular system capacity where significant frequency reuse is employed--approximately where the subscriber base exceeds about 15,000--all costs are variable, on a per-site and per-equipped-channel basis.

Analysis

Figure 1 has been drawn to show the general nature of cellular mobile system investment per subscriber as the number of subscribers in the system grows. Initially, the system is constructed so as to obtain the needed geographical coverage and radio channels are added to add capacity as required to the fixed number of sites. When the capacity of the whole set of 395 "voice" channels becomes exhausted, at about 10,000 subscribers (25 subscribers/channel x 395 channels = 9,875) some frequency reuse becomes necessary. Over this range of system capacity, the large fixed costs associated with the large number of sites, the mobile telephone switching office, separate office space, etc., make the investment per subscriber fall rapidly as the number of subscribers grows.

We define the following terms:

I_{fs} = Fixed investment per subscriber = ((Fixed cost per site) x (Number of sites))/(Number of subscribers)

I_{vs} = Variable investment per subscriber = (Per-equipped-channel cost)/(Number of subscribers per channel)

$I_s = I_{fs} + I_{vs}$ = Total investment per subscriber

Conceptually, investment cost per subscriber behaves in a "lumpy" fashion as new sites are added when capacity at existing sites runs out. In Figure 1 we show what would happen just before and just after the point that all sites run out of capacity, assuming that they all run out at the same time and that the number of sites is doubled each time that happens. Thus, we see I_{fs} doubling at the point that the number of sites needs to be doubled. To a reasonable approximation I_{vs} grows linearly with the capacity needed in terms of the number of subscribers. At the point the number of sites needs to be doubled, the existing number of equipped channels can be redistributed amongst the sites, to a rough approximation.

Of course, in reality, the system will not grow in the fashion described above. Instead, cells are split and added as needed, where needed, and, on average, each cell will be equipped with somewhere between the maximum number of channels that can be used in a cell and half that number. Accordingly, this is illustrated in Figure 1 by the smoothed flat horizontal line labeled I_s. The labeling of the fixed and variable costs on the Figure may be a little confusing; however, the thing to keep in mind is that the "smoothed" investment per subscriber is made up of a "fixed" cost and a "variable" cost. The former is associated with the fixed cost per site and the number of sites while the latter is associated with the capacity needed (number of subscribers) and, therefore, strongly related to the number of equipped radio channels required. One then finds that all costs in a cellular system are "variable" even though some change in a rather "lumpy" fashion, but only on a site-by-site basis and not for the system as a whole. I_s stays relatively constant as capacity is increased through successive stages of frequency reuse. Thus, over a very large range of capacity, cellular systems do not exhibit economies of scale; marginal costs can be approximated by average costs.

Adding a small amount of spectrum, say 1 MHz, delays the needed addition of new sites and reduces the number of equipped radio channels that would otherwise be needed for supporting a given level of capacity (subscriber base). Note that with 395 voice channels available and the assumed "average" configuration of the 7 cell, 3 sectors per cell, system, there is available a maximum of 18.810 channels to trunk together in each of the 21 sectors. On average, however, cell sites will not have the maximum number of channels as the system capacity expands.

If the number of sites doubles as capacity doubles, as postulated above, then the number of channels per sector would range from about 10.983 to 18.810 (10.983 instead of 18.810/2 = 9.405 due to the lower trunking efficiencies). At the point the system capacity runs out and the number of sites needs to be doubled the capacity per site is halved. Under the 5% erlang B blocking probability, assumed throughout this paper, 10.983 channels will carry half the traffic that 18.810 channels will carry. This takes place as new sites are added and reconfigurations are needed to match capacity requirements to traffic requirements. Thus, on average, the sectors (and, therefore, trunk groups) would be equipped with about (10.983 + 18.810)/2 = 14.897 channels, for 44.691 channels per cell.

Continuing with an assumed blocking probability of 5% we can compute the added capacity that an added one MHz of spectrum provides. For the "average" loaded site each trunk group has 14.897 channels, as computed above. When one MHz is added this provides 1/0.06 = 16.67 channels to add to the 395 channels. When these channels are distributed over the 21 groups of channels we get 16.67/21 = 0.794 channel that can be added to each trunk group. The maximum number of channels in a trunk group now becomes 18.810 + 0.794 = 19.604. Repeating the above calculations to obtain the average number of channels in the enlarged trunk groups we find that it takes 11.414 channels to carry half the traffic that 19.604 channels carry just before cell addition. As before, we estimate that the average number of equipped radio channels is approximately given by the average of these two numbers or (11.414 + 19.604)/2 = 15.509.

Now a couple of simple calculations can be made that will allow us to compute the value of the added spectrum once the fixed and variable costs of cellular systems are estimated. The needed quantities are the carried traffic (erlangs or E) per sector or trunk group before and after the added MHz of spectrum. For the 5% blocking (P05) 14.897 channels will carry 10.54 E and 15.509 channels will carry 11.09 E. Thus, the increased traffic carrying capacity per site is ((11.09 - 10.540)/(10.54)) x (100) = 5.2%.

As soon as the fixed per-site and variable per-equipped-channel costs are tied down we can compute the value of the added one MHz to a cellular system having only 25 MHz to work with. The value of the added spectrum comes about because the number of sites need not be as great, for a given total capacity requirement, and, therefore, the number of subscribers.

Actually, an additional savings is realizable because the capacity of each equipped channel goes up because of increased trunking efficiency. This amounts to over a 1% increase so that, for a given level of total capacity, the number of needed equipped channels is reduced by this amount at each site. Thus, the per-equipped-channel costs at each site should be reduced by this amount, in addition to the savings from the reduction in the number of required sites. At this point we need to estimate some costs.

We were able to estimate the fixed cost per site from June 1982 FCC 401 forms for the initial 24-site Los Angeles AMPS system. Per site costs were given along with the number of equipped channels. By fitting a curve to this data we were able to estimate the per-site investment cost per equipped channel, I_c, as follows:

$I_c = 4.3 + 483/N$ $k/channel, where N = Number of channels.

To the above we estimate that one should add $2k per cell-site channel for a microwave link to the mobile telephone switching office (MTSO). Further, we estimate that one can add $3k per channel for the MTSO switch. Coupled with these add-on choices we believe one can largely ignore the rather small relative investment contributions resulting from the general office space needs and the fixed costs associated with the MTSO. This, assumes, of course, as stated above, that we are discussing only systems that are employing significant frequency reuse, where the fixed costs of the multiplicity of sites dominates the other fixed costs of the cellular system. After these additions, our final estimate for the total investment cost per equipped channel in a cellular system becomes:

$I_{ct} = 9.3 + 483/N$ $k/channel, where N = No. of channels/site.

Note that our estimate for the fixed cost of each cellular system site is $483k. Finally, we can estimate the cellular system investment saving that results from the added one MHz of spectrum. As indicated above, the primary savings come about from two factors: (1) fewer sites are needed and (2) fewer equipped radio channels are needed at any given level of system capacity. Fewer sites are needed since each site can carry more traffic and fewer equipped channels are needed since each channel carries more traffic when more channels are available due to increased trunking efficiency.

For simplicity consider a 100-site cellular system. With the additional MHz of spectrum each site can carry 5.2% more traffic, and, therefore, to a close approximation, 5.2% fewer sites are needed. In this otherwise 100-site system a saving of 5.2 sites saves an investment of (5.2 sites) x ($483k/site) = $2.51M due to the fixed per-site cost savings. The further savings from the reduced number of equipped radio channels can now be computed.

The 100-site system has (14.897 channels/sector) x (3 sectors/site) x (100 sites) = 4,469 equipped radio channels. The system with the added spectrum only needs 100 - 5.2 = 94.8 sites to serve the same subscriber base. The number of equipped radio channels becomes (15.509 channels/sector) x (3 sectors/site) x (94.8 sites) = 4,411. Thus, the added spectrum saves the cellular system 4,469 - 4,411 = 58 equipped radio channels. Recall, that we estimate that the total per-equipped-radio-channel investment cost is $9.3k. This results in a savings of (58 radio channels) x ($9.3k/channel) = $0.54M. Finally, we obtain an estimated total investment saving of $2.51M + $0.54M = $3.05M. This is the savings that would be realized from one MHz of additional spectrum being supplied to an existing 100-site cellular system.

Over the range of model applicability (significant frequency reuse, or, equivalently, more than about 15,000 subscribers), the above investment savings (or value of spectrum estimates) can be scaled linearly with the number of sites in the system and the amount of added spectrum, at least up to several MHz. As the added spectrum goes above the one MHz considered the results become more and more conservative as not all of the increased trunking efficiency, associated with the larger trunk groups will be included. We define the following terms:

V_S = Value of spectrum to a cellular system,

N_S = Number of sites in a cellular system, and

S = Amount of additional spectrum, in MHz.

The results of the model then gives, in formula form:

V_S = 0.0305 x (No. of sites) x (Spectrum amount, MHz) $M

= 0.0305 x N_S x S (in MHz) $M.

For example, the Los Angeles PacTel system presently has around 100 sites which suggests that one MHz of additional spectrum would be worth about (100) x ($0.0305M) = $3.05M to this cellular system provider. For a 10-channel 900 MHz system, utilizing 0.250 MHz of spectrum, the value is $762.5k. A 10-channel 800 MHz trunked system uses twice as much spectrum so that the value of its 0.5 MHz of spectrum to such a cellular system provider would be twice the latter amount, or $1.525M, etc.

If one knows both the number of sites in a cellular system and the number of subscribers a much better estimate of the value of spectrum to the cellular system provider can be made. One can more closely compute the average number of equipped channels per site than was done by assuming/estimating the number of cells in each use of the available channels and the number of associated sectors per cell. This approach further required estimating the average number of channels in each cell.

The better approach, when the number of subscribers is known, is to use the reasonably good "rule of thumb" that each equipped radio channel can handle 25 subscribers. (In the larger systems 30 subscribers per channel may be a better estimate.) This allows the number of equipped channels to be computed from the number of served subscribers. Finally, the number of equipped channels per site can be calculated.

For example, for the PacTel system having about 100 sites and 170k subscribers, we get 1.7k subscribers per site and 1.7k/25 = 68 channels per site. Recall that, for the chosen configuration, the maximum number of channels per site is about 57 and we used an average of less than 45. For the 4-cell, 6 sectors/cell, reuse pattern the maximum number of channels per cell is 98 to 99. This indicates that the PacTel system is using a lot of this maximum density pattern; in fact, significantly more than estimated above. However, the 4-cell, 6 sector/cell reuse pattern divides the available channels into 24 trunk groups, compared with the 21 groups we worked with. Thus, differences in trunking efficiency gains between the two choices may not be large. And, for reasonable increases in amount of spectrum, savings due to increased trunking efficiency are small compared with savings due to the reduction in the number of sites. One can easily see that there are a number of approaches to getting useful results, with some being better than others, depending upon the purpose at hand.

It is very likely that the fixed cost per site is higher than what we used. If the 1982 number of $483k per site was correct then, this value should likely be grown by at least a very conservative compounded annual growth rate of 5%. (As cell sizes diminish the latitude in selecting the locations for sites is reduced, and, therefore, the difficulty, and expense, in acquiring them goes up. This just suggests that even the 5% growth factor may be much too low.) This would increase the $483k to $647k per site. One would have to go through the calculations, but it is easy to see that the added one MHz of spectrum would be worth much more than what was calculated above. This can be easily scaled. One finds that the value of an added MHz of spectrum to a 100-site cellular system increases from $3.05M to $3.90M. Using the above later estimate that each PacTel site is equipped with about 68 channels and the higher fixed cost per site of $647k, the investment per site for this system may be on the order of $1.3M, and possibly more due to alleged difficulty in acquiring sites as cellular systems grow.

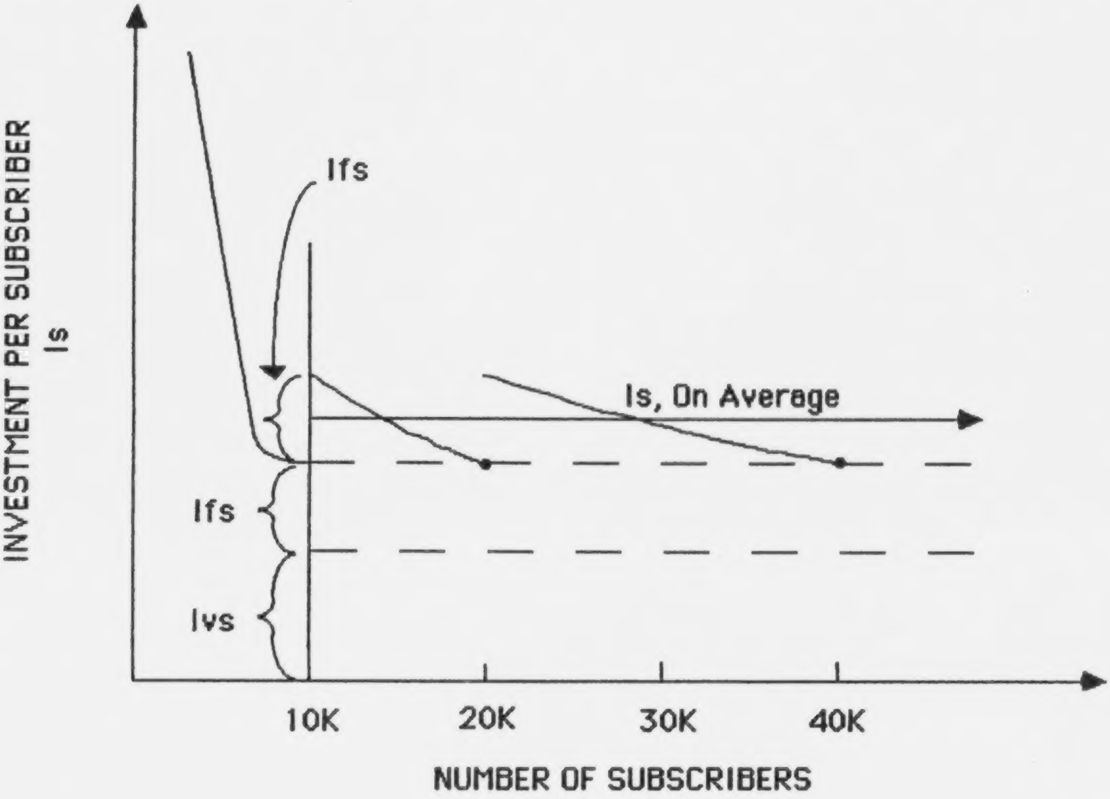


Figure 1: CELLULAR MOBILE SYSTEM INVESTMENT PER SUBSCRIBER VERSUS NUMBER OF SUBSCRIBERS

Privatization of NTT in Liberalized
Japanese Telecommunications Market

Submitted to the 16th Annual
Telecommunications Policy Research Conference
Airlie House, Airlie, Virginia
October 30 - November 1, 1988

Koichiro HAYASHI and Hideaki TODA
Nippon Telegraph and Telephone Coporation (NTT)

* The views expressed in this work do not necessarily represent the official views of NTT.

1. Introduction

Three and a half years have passed since the liberalization of the telecommunications market in Japan and privatization of the Nippon Telegraph and Telephone Public Corporation in 1985.

Because full competition has not yet been achieved in this field, the pre-scheduled review of the Telecommunications Business Law, which had prescribed the liberalized market structure, has been postponed until 1990. In that year, the review of the NTT Corporation Law, will be completed.

In this context, it is too early to judge whether the liberalization of the telecommunications industry and the privatization of NTT has been successful or not.

However, eight or nine years have already elapsed since the preliminary stage of liberalization and the heated debates at meetings of the ad hoc commission for administrative reform (RINCHO). And in some specific business fields, competition has been successfully introduced and the customers are taking advantage of it.

This paper focuses on the performance of NTT after privatization, but it inevitably relates also to the liberalized telecommunications market structure and market performance.

2. Framework for Telecommunications Business

One characteristic of Japan's telecommunications reform is that all fields within telecommunications - even local exchange services- have been opened to competition. I believe this contrasts sharply with the reforms under United States legislation, which retain treatment of local services as a natural monopoly.

Another feature of the Japanese system is the clear distinction between Type I and Type II carriers. Type I carriers are specified by the telecommunications business law as those providing telecommunications services via their own telecommunicatios circuit facilities. On the contrary, Type II carriers don't have their own circuit facilities but lease them from Type I carriers.

As Type II activities mainly involve adding value to networks, Type II carriers are considered to be VAC or VAN carriers. Type II carriers are further divided into two types. One is Special Type II, which provides large-scale nationwide services and/or international services. The other one is General Type II. And, an additional category called International Special Type II was created consisting of the international VAN services between United States and Japan following the reaching of a bilateral agreement.

The degree and content of government regulation vary by category. Type I carriers must submit to the severest regulation, while General Type II enjoy a fair degree of freedom to go about their business activities. This dichotomy of Type I and Type II carriers is unique to Japan.

Table 1. Industry Framework

	Type I	Type II	
Business Activities	Provision of telecommunications services through own telecommunications circuits and facilities	Provision of telecommunications services through telecommunication circuits leased from Type I carreirs	
		Large-scale nationwide and/or international services	Other services
Government Regulation			
a. Start-up of services	Permission from MPT required	Registration with MPT	Notification to MPT required
b. Rates and Charges	Authorization from MPT required	Notification to MPT required	Unregulated
Foreign Capital Participation	Proportion of total shares held by foreign nationals limited to one-third	Unregulated	Unregulated

3. New Entrants as Type I Carriers

As of August 1, 1988, 39 carriers (other than NTT and KDD) have obtained permission from the Ministry of Posts and Telecommunications (hereafter MPT) to operate as Type I telecommunications carriers. Among them, all of the land-based carriers and local carriers, as well as most of the paging service providers have already begun operation.

As for the Type I telecom carriers, the terrestrial long-distance interexchange carriers are most actively pursuing their business. But I would like to return to this point later.

Two satellite carriers are preparing to launch their satellites in partnership with U.S. satellite manufacturers. In Japan, hitherto, only limited satellite services have been provided. A key to their success is that customers have shown greater interest in satellite communications services.

While in the United States no competition is allowed in the local telephone market, in Tokyo, a regional carrier, called TTNet, is already providing both private leased circuit services and public switched telephone services. TTNet uses existing equipment and facilities owned by Tokyo Electric Power Co., a parent company. Like the three national long-distance carriers, it plans to provide services at rates about 20% lower than those of NTT.

Two land-based cellular mobile radio phone companies have obtained permission to operate - one in western Japan and eastern Japan respectively. They will start providing service in late

1988 or thereafter. It is expected that the Japanese cellular radio phone market will also grow rapidly. Another unique company called Tokyo Bay Marinet, is focusing on the provision of ship telephone and portable, private-phone service.

New radio paging carriers are rapidly growing not only in number but in terms of number of customers and service areas. MPT has adopted a policy of licensing only one new radio paging carrier per prefecture to compete with NTT. The number of radio paging carriers has reached 24, but will continue to grow.

Table 2. Type 1 Carriers

(Excluding NTT and KDD)

Classification			Licensed Carriers	Operating Carriers
Domestic	Leased Lines and Telephone	Land-based	3	3
		Satellite	2	0
		Local	4	4
	Cellular Telephone		3	0
	Paging Service		24	19
	Other Services		1	1
International (Leased Lines and Telephone)			2	0
Total			39	27

As of Aug. 1, 1988

All the three land-based carriers (usually called NCC for New Common Carriers) now provide private leased circuit and public switched telephone services in three major cities, Tokyo, Nagoya, and Osaka. And during a second phase, two of the three plan to expand their services to other geographical areas, - Fukuoka in southern Japan, and Sendai or Morioka in northern Japan.

All of these areas are densely populated, and have heavy telephone traffic. With their low-cost services usually 15% to 20% less expensive than similar NTT services, these three carriers have already won a large number of customers. In the case of telephone service, NCCs have won about 3 million contracts.

Each of these new carriers has financially strong foundations. DDI is a joint venture of the companies Kyocera, Secom, Ushio Electric Co., and others, - companies epitomizing the entrepreneurial spirit in the Japanese corporate world. And Japan Telecom has a close relationship with Japanese Railways and is going to merge with Tetsudo Tsushin, which is the sole provider of service (including ticket reservations) to the whole JR group.

The last one, Teleway Japan, is primarily owned by the Japan Highway Public Corporation, and Toyota Automobile Manufacturing Co. also strongly supports it in terms of both equity and sales promotion.

Table 3. Type I Carriers

Land-Based Carriers (Leased Lines and Telephone)

	Start of operation	Number of customer constracts	Rates compared to NTT's
Daini Denden Inc. (DDI)	Leased lines Oct. '86 Telephone Sept. '87	* Leased lines (# between Tokyo/ Nagoya/Osaka) 1,000 * Telephones 3,000,000	* Leased lines Analog: 8% lower Digital: 13% lower * Telephone max. 30% lower
Japan Telecom	Leased Lines Aug. '86 Telephone Sept. '87		
Teleway Japan	Leased lines Nov. '86 Telephone Sept. '87		

Let me briefly explain the present status of Japan's international telecommunications market. Three land-based Type I carriers have begun operation in the domestic telecommunications market. In contrast, in the international field, KDD still retains its monopolistic status. However, two international NCCs have obtained permission to enter the market and they will begin their leased line services next year.

ITJ is a joint venture among large trading companies and Matsushita Electric. IDC is a consortium of C. Itoh - a leading trading company of Japan, C&W Corporation, Toyota, Pacific Telesis International, and others. As the trading companies lack extensive telecom business experience, these new companies need technical assistance. KDD is assisting ITJ, while NTT International is serving as technical adviser to IDC.

Table 4. Type I Carriers (Cont.)

Internatinal Carreirs

	Services Offering	Start of Operation	
International Telecom Japan (ITJ)	Leased lines Telephone	Leased lines Telephone	Apr. '89 Feb. '90
International Digital Communications (IDC)	Leased lines Telephone	Leased lines Telephone	May '89 Oct. '89

4. Type II Carriers

There are approximately 600 Type II telecom carriers, now offering various services. About 70% of them offer data transmission, and about 30% provide voice services. Among the carriers, 46 are resellers.

Although the total number of providers is rather small, it is worth noting the existence of video offerings, even among General Type II carriers. I am sorry this information is not included among these figures, but about a half of the General Type II carriers emerged from Computer Service Bureaus. This implies that they have adequate experience in the information-related industry and stable financial resources. For that reason, we foresee the phenomenal expansion of this business in the near future.

Special Type II carriers are rather large concerns at the leading-edge of this market. A great many of these emerged from main-framers and from service bureaus. There are of course completely independent service providers, but the number is small.

Of the 20 carriers in the Special Type II category, 11 are considered International Special Type II carriers. As I have already mentioned, this category was established in accordance with the mutual agreement regarding international VAN services between the U.S. and Japan last year.

Table 5. Type II Carriers

Classification		No. of Carriers	Service Offerings (Multiple Offerings)	
Domestic	General Type II	567	Voice (Reseller	165 46)
			Video	84
			Data	400
			Voice/Video/Data	70
			Total	719
	Special Type II	20	Voice/Video/Data	
International (belongs to Special Type II)		11	Data/E-Mail/Database	

(As of August 1, 1988)

5. Estimation of Market Shares

I have talked so far about the number and names of the new entrants. But how about their market shares?

It is fairly difficult to say anything concrete about market share, because there are neither official statistics nor reliable statements. We have to gather information by ourselves.

One information source we have is the number of customer contracts for telephone service which is reported in the media. According to this data, the NCCs have already won 3 million contracts - a 7% market share. But market share as measured by number of contracts is not equal to market share as measured in terms of real traffic. And there is supposed to be a large number of customers who have made contracts with all 3 land-based carriers.

But at the same time, it is very important to note that their service areas are limited to the main cities between Tokyo, Nagoya and Osaka, while NTT provides nationwide telephone services. And, with that latter factor, the number of their contracts should increase rapidly together with service areas.

The contract data for leased lines are somewhat closely correlated with market share data even though differences in tariffs exist according to the types of circuits. In this field, NCCs have already established about a 15% market share. This figure is remarkable, because not even two years have elapsed since their start of operations.

In some business fields, competition is stronger, and market structure is evolving rapidly. For example, in paging services, NCCs have reportedly acquired a 10% market share in the only half year since their entry. Most notably, in Kumamoto prefecture on Kyushu Island, one of the new paging companies has actually exceeded NTT in terms of number of customers, and its market share is estimated at approximately 60%.

Last but not least, the Customer Premise Equipment (CPE) market is the most competitive field. This has been opened to competition in 1985 simultaneouly with market liberalization and NTT privatization. Customers now can choose any kind of CPEs including main telephone as they like. And there are so many CPE vendors, some of which are traditional communication equipment manufacturers but others entering are household electronic manufacturers. NTT's sales share in this field is estimated to be about 25% according to my investigation, though this figure does not include the revenues for CPE rentals, which are provided only by NTT.

Table 6. Market Shares (estimated)

Telephone	No. of customer contracts	% share
NTT	47.5 million	93
NCCs*	3.5 million	7
Leased lines (between Tokyo/Nagoya/Osaka)	No. of circuits	% share
NTT	5,700	85
NCCs*	1,000	15
Paging Services	No. of customer contracts	% share
NTT	2.7 million	90
NCCs*	0.3 million	10
Customer Premises Equipment	Sales Revenue	% share
NTT	¥176 billion	25
Other vendors	appr. ¥530 billion	75

* NCCs: New Common Carriers (including all kinds of Type I carriers).

6. Liberalization and NTT

Liberalization of the telecommunications market in Japan has meant two things for NTT. It has freed us from the protective skirts of the Government, increasing competition and removing the government's financial guarantees. At the same time, it's increased our flexibility and given us the opportunity to be creative and innovative in our business activities.

Until the telecom restructuring of 1985, the old NTT had been "protected" by the Government. This protection took two forms. We had been assured of a monopoly in our area of service, and we didn't have to pay any taxes. On the other hand, we were also subject to severe regulation. Concretely, this meant our overall plans for cash flow, investment, personnel and other areas of management required approval by the Diet -- that is, the Japanese Parliament -- as part of the entire governmental budget. In virtually all respects, we were a governmental institution.

Restrictions in a real sense went beyond those I just mentioned and into the realm of internal mechanisms. Responsibility for income and expenditures, as well as service management, was often not clear. The emphasis in personnel management was on who to place in what administrative role. A demerit system was employed where every time a mistake occurred a certain number of "points" was deducted on the employee's personnel record.

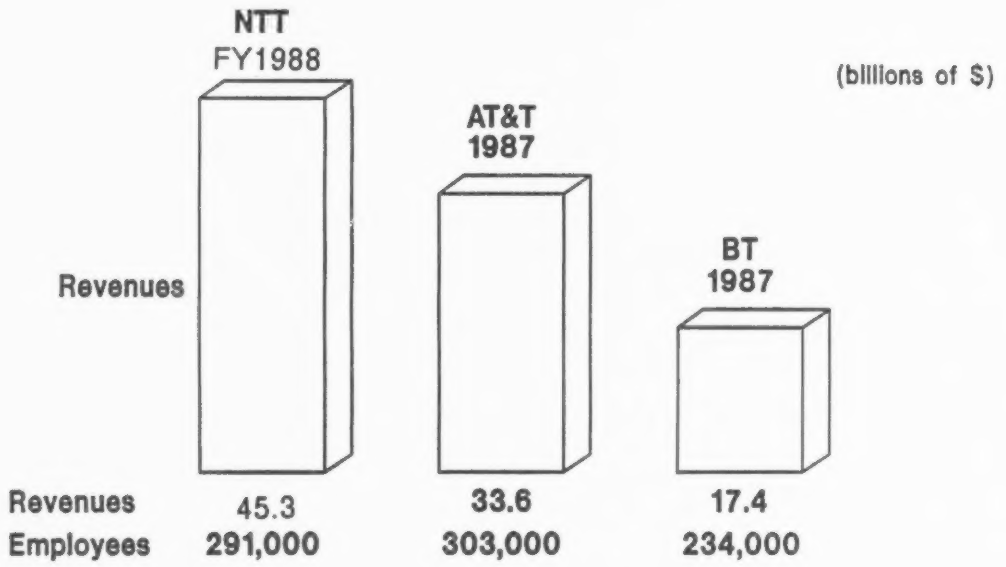
What I'm trying to show here is that we had "administrative control" rather than "enterprise management". Employees felt their responsibilities were fulfilled if they did what was specified in the specified way. A "budget-centered" psychology ensured that what was budgeted would be executed, even if there was no more need for such an action. On the reverse side of this coin, even if the need for some new measure might arise, there had been no way to execute it if there was no budget to do so. The rigidity of corporate mechanisms made it almost impossible to instill the creativity and ingenuity needed to assure better service and increased profits.

7. NTT's Performance after Privatization

As you may not be acquainted with NTT, let me note that NTT is a leading domestic common carrier in Japan. Our corporation has performed well since its privatization and, in fact, were you to look at the results for the fiscal year ending March 31, 1988, you would find that NTT has become the largest of the world's common carriers both in terms of revenues and number of employees. (See Figure 1.)

Although this may be partly ascribed to the significant appreciation of the Japanese yen within the last 3 or 4 years, we are very much satisfied with these results and very proud of our company's performance.

Figure 1. **MAJOR WORLD CARRIERS**



NTT has been making every effort to improve its management both by increasing revenues and by reducing expenditures. As a result, it has been able to repeat its good performance year after year. Results for fiscal 1988, were particularly outstanding. Pre-tax income in FY1988 is estimated at about 450 billion yen - an increase of 25% over the year earlier.

In addition to having accomplished the goals it initially set for itself, NTT has reduced its tariffs three times since privatization. The first time was in 1986 when it reduced Saturday rates 40% - the same discount that is applied to Sundays and Holidays. The second time was in 1987, when leased line tariffs were reduced to remain competitive with those offered by the NCCs. The most recent rate reduction - in February this year - focused on a 10% cut in rates for long distance calls, and on a substantative improvement in rates for calls from remote islands.

As for long distance call rates, NTT has reduced its long distance tariffs three times since 1984. Before 1981, rates for domestic calls made over the longest possible distance were notoriously 72 times as high as rates for local calls. That ratio has since been halved and rates for calls made over the longest distance are now 36 times as high as for local calls. But, we still believe those long distance rates are too high relative to the cost of providing them, and we are ready to reduce them further.

NTT's network facilities have traditionally depended on analog systems. But we are now rapidly moving to introduce digital facilities, partly to meet the requirements of customers seeking to use ID capabilities and other advanced capabilities and partly to reduce network costs by taking advantage of new technologies. The ratio of digitization was 15% at the end of fiscal 1988, and it is expected to exceed 40% by the end of fiscal 1990.

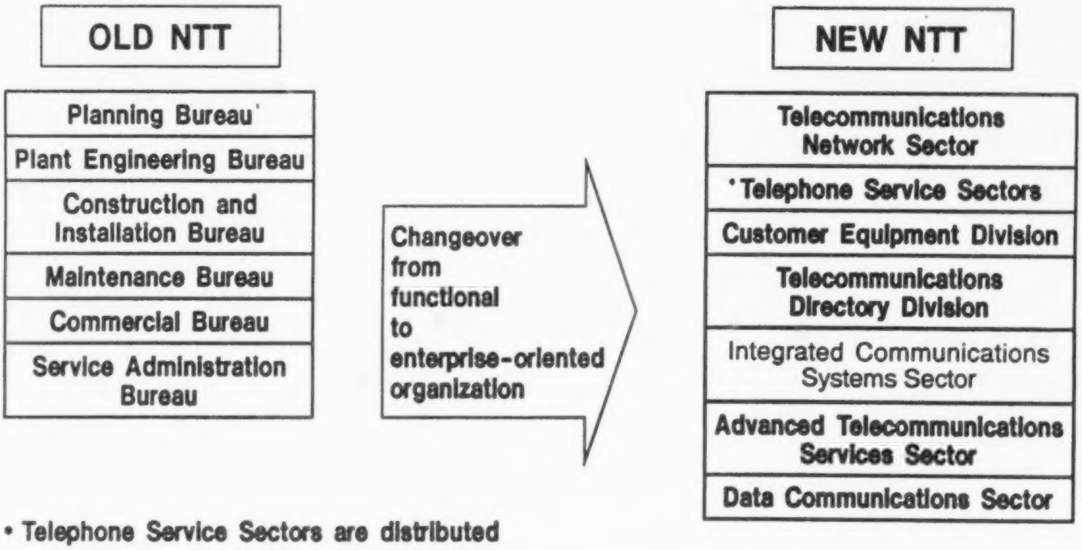
8. Restructuring of Organization

With privatization as impetus, NTT was eager to change its organization and restyle its culture.

But, given the huge size of NTT, it will take several years to fully realize these changes in the NTT's organization. Three years and a half have elapsed since NTT's privatization and its reorganization process is still underway.

The first step had been the changeover from a bureaucratic, functional organization to an enterprise organization along profit center lines. This is being achieved with the introduction of a profit center system.

Figure 2



RESTRUCTURING OF ORGANIZATION ALONG PROFIT CENTER LINES

OLD NTT
Planning Bureau
Plant Engineering Bureau
Construction and Installation Bureau
Maintenance Bureau
Commercial Bureau
Service Administration Bureau
Changeover from functional to enterprise-oriented organization
NEW NTT
Telecommunications Network Sector
• Telephone Service Sectors
Customer Equipment Division
Telecommunications Directory Division
Integrated Communications Systems Sector
Advanced Telecommunications Services Sector
Data Communications Sector

• Telephone Service Sectors are distributed corresponding to distinct national regions.

We have established five profit centers 1. Telecommunications Network Sector, 2. Telephone Service Sector, 3. Integrated Communications Services Sector, 4. Advanced Telecommunications Services Sector and 5. Data Communications Services Sector. And the telephone service sectors are again organized to correspond to eleven national regions.

The objective of this new organization is to cope with competition, of course, and to respond more quickly to changing cutomer needs. Gradually, it is being realized that we should be paying more attention to our major customers.

There are two reasons; the first being that they are the customers that provide us with significant revenues. In the case of leased circuit service, the 200 largest customers account for half of NTT's total leased circuit revenues.

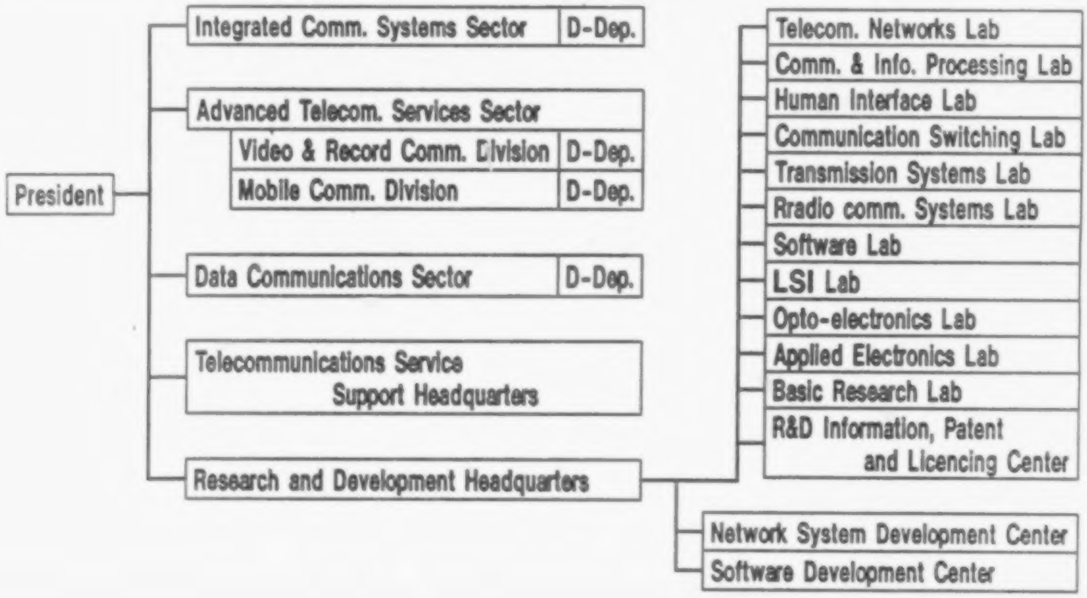
Secondly, major customers represent the leaders of technical innovation in the telecommunications isdustry. It is essential for NTT to satisfy customer needs to realize technological advances.

To give more attention to our special customers, we have introduced a system of account management teams. A single AM team represents each of NTT's service sectors and responds to needs of specific clients. Major customers can easily access NTT and the far reaches of its organization via a single AM team.

To cope with our changing environment, we are also finding it necessary to strengthen our research and development functions. We are reorganizing our laboratories to be more customer-oriented and responsive. I would like to remind you, too, that NTT has no manufacturing arm just as in the case of most PTTs. Of course, we are in very close contact with telecommunication, equipment vendors but our main domain is the provision of telecom services not hardware. In this context, it is most important for us to monitor and satisfy customer needs. The research and development organization was reorganized into eleven labs, one center, and two development centers. Simultaneously, some development functions closely relating to market needs were transferred to the pertinent profit centers.

Figure 3

NTT R&D ORGANIZATION


President
Integrated Comm. Systems Sector
D-Dep.
Advanced Telecom. Services Sector
Video & Record Comm. Division
D-Dep.
Mobile Comm. Division
D-Dep.
Data Communications Sector
D-Dep.
Telecommunications Service Support Headquarters
Research and Development Headquarters
Telecom. Networks Lab
Comm. & Info. Processing Lab
Human Interface Lab
Communication Switching Lab
Transmission Systems Lab
Rradio comm. Systems Lab
Software Lab
LSI Lab
Opto-electronics Lab
Applied Electronics Lab
Basic Research Lab
R&D Information, Patent and Licencing Center
Network System Development Center
Software Development Center

With these R&D changes, career patterns for R&D staff were affected. Under the former system of lifetime employment, it had been very exceptional for anyone involved in R&D to be given exposure to the business front over the course of their career. Under this reorganization though, some R&D staff are now being assigned to our development centers or business sectors to test the applicability and marketability of their inventions and, in turn, to feedback that information to the next phase of R&D.

The last task we face is that of reinforcing our organization inside the telephone office. NTT has more than 1300 telephone offices serving as the company's interface with customers. Having been a part of government, NTT suffers the legacy of being overly centralized in terms of authority and human resources. We are trying to undo that legacy by decentralizing it. In reshaping the company's organization, we are in many cases eliminating intermediate headquarters by integrating district and regional headquarters.

With its privatization, NTT was also given the right to freely establish subsidiaries and affiliated companies. Since 1985, the number of such new companies reached 180. In the case of one third of NTT's subsidiaries and affiliates, the motivation for establishment was utilization of NTT-owned technology. We find that about 30% of these companies are software and information processing companies and another 30% are VAN/LAN and database service providers.

9. Changing Corporate Culture

In pace with these phenomenal changes in NTT's activities, NTT's corporate culture has been changing gradually. People had been complaining about NTT being conservative and bureaucratic, "slow'n steady" and all that. Looking at the company today, though, you would find an increasing sense of dynamism. It is becoming an enterprise willing to seek and take on new challenges and an enterprise that is responsive to customer needs.

NTT was awarded a Triple-A rating by both Standard & Poor's and Moody's and is now ranked No. 1 in the Japanese recruitment market. A positive reputation is emerging because of this increasingly upbeat corporate culture.

I would like to remind you that this is not a natural stream of events but, rather, the result of deliberate effort. There are several movements underway to change corporate culture.

"Orange Line" is the name of a system to give users of NTT services a voice. It encourages NTT employees to listen to customer needs and to respond to them quickly. "MI" which stands for Management Innovation is a second movement underway affecting every facet within NTT, while "ASK" is yet another movement which promotes quality control within NTT.

ASK is an abbreviation for the words in Japanese - Anzen, Sokuoh and Koritsu. The translation would be Safety, Quick Response and Efficiency. This concept relates to a philosophy of Dr. Shinto who has been engaged in the shipbuilding industry for a long time, and who is famous not only as the designer of the so-called "Shinto type of oil tanker", but also as a successful executive in IHI. During this experience, he had realized that safety and efficiency are two sides of the same coin. After Dr. Shinto undertook new responsibilities as the President of NTT, about four years prior to its privatization, he has been making every effort to introduce a kind of TQC System in NTT.

A new program, we are now moving to introduce is "TPU" short for Total Power Up. TPU focuses on increasing revenues and reducing expenditures by making use of the employee ideas. It is a bottom-up movement. And, in conjunction with the strong top-down leadership of Dr. Shinto, I believe this movement can enhance our corporate culture.

10. Some Remaining Problems

Let us come back now to the topic of the telecommunications industry itself. The Telecommunications Business Law uniquely prescribes for that law to be reviewed within 3 years of its enactment, and for any necessary amendment to be taken.

In accordance with this provision, MPT held a hearing with the various parties concerned and, of course, NTT was invited to the hearing. NTT issued the following statements at the hearing.

Review of the Telecommunications Business Law should be divided into two phases. In the long-term, regulation itself should be abolished, and the convergence of telecommunications and broadcasting may give rise to other legal problems - perhaps necessitating the review of the Broadcasting Law and the possible enactment of an entirely new law to cover both telecommunications and broadcasting. These two items are likely to be long-term subjects of review.

But in the short-term, there are three telecommunication matters demanding urgent attention. The first is the relaxation of licensing requirements, including the relaxation of tariff authorization. Secondly, there is the spin-off of NTT's Data Communications Sector. And, third is the question faced by the MPT of whether to allow NTT to provide leased lines to CATV suppliers.

Unfortunately, the balance of public opinion remains strongly in favor of the NCCs. Of course, they are not against NTT but they encourage us to behave as a dominant carrier.

Furthermore, a minority opinion insists upon NTT's divestiture.

Under these circumstances, the review of the law was postponed. MPT would like to review it again in accordance with the review of NTT Corporation Law that is scheduled to take place a year and a half from now. For now, only one item that was considered - the spin-off of the DCS - has been realized.

And NTT is now faced with a very curious situation, in which the stronger it becomes in the market the stronger is the voice calling for the divestiture of NTT. Meanwhile the weaker NTT becomes, the more strongly public opinion offers NTT its encouragement. But curiously no one has decided what market share is appropriate for the giant NTT to hold. Nobody knows how effectively universal service can be maintained if NTT is divested.

It may seem very strange to people in the United States, for there has not been sufficient data or analysis with respect to regional or sectoral cross-subsidization. In this context, we have only just become aware of the market nature of the telecommunications industry since its liberalization.

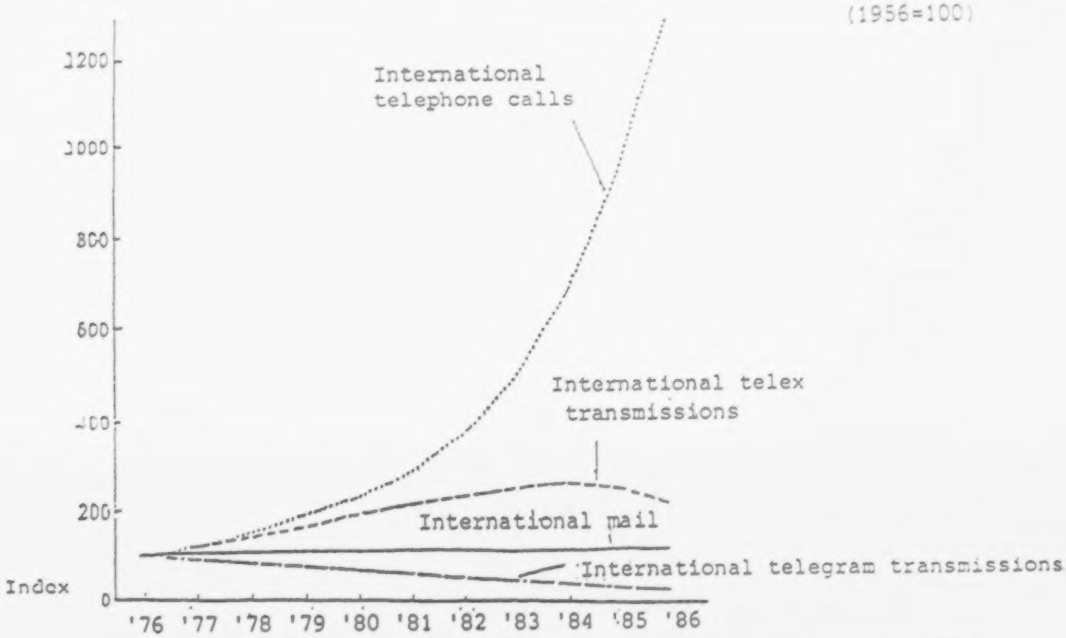
11. Toward Expansion of the Market

Although the review of the Telecommunications Business Law has been postponed, we still can examine figures that illuminate market trends.

Figure 4 shows the change in per-person use of the respective principlal domestic media in Japan. You can easily understand how rapidly the number of telephone calls per-person has grown since 1980. This phenomenon was not caused by the liberalization itself, because the market-opening was in 1985, as I have already mentioned, although liberalization may have played a part in promoting the rapid growth.

Figure 4

Trends In International Telecommunications



Compiled from KDD's and the Ministry of Posts and Telecommunications' data

Note: 1. International mail includes total dispatches and arrivals.

2. International telegrams, telexes, and telephone calls include total dispatches, arrivals, and relays.

Next, let's take a quick glimpse at the size of Japan's Information/Communication Market in terms of annual sales.

In 1985, annual revenues of Type I carriers were estimated at 5.3 trillion yen. This market is fairly large when compared with other markets such as broadcasting, even though the market size is approximately one fourth of that in the U.S.

The combined annual sales of Type II carriers are estimated at 520 billion yen. This figure comes from the publication, "A Future Vision of Advanced Telecom Services" prepared by the Telecommnications Council and published by MPT in 1987. But as shown in the footnote, these figures should be seen as representing the information services industry as a whole, and the estimated annual sales of Type II carriers as defined in a strict sense was considerably less than 50 billion yen - somewhat less than one tenth of that 520 billion yen figure. This may be verified by the fact that NTT's packet switching service revenue was only about 10 million yen in 1985.

Compared with these annual revenue figures, you may be surprised to see the proportion of sales that are included under the category of manufacturing of communications equipment. Even though this figure includes the manufacture and lease of computers, you may find good market opportunities in this field.

At any rate, the information/communication related industry has been developing soundly and rapidly these past several years, and it is expected to continue to grow.

Policy-making should be based upon this confidence, and focus on how to expand the market. NTT's divestiture problem should also be judged from this point of view.

Table 7. Size Of Japan's Information Communications Market in Terms Of Annual Sales

(100 million yen)

	1981	1982	1983	1984	1985
Telecommunications					
Type I	42,625	44,611	49,915	48,854	53,074
Type II					5,200 *
Broadcasting	12,236	14,550	15,499	16,290	16,240
Rural telephone with broadcasting function	182	187	170	174	187
Telecommunications-related Sectors					
Manufacture of communication equipment	168,589	188,225	228,346	278,241	303,002 **
Lease of communications equipment	7,209	7,999	11,154	12,899	14,234 **
Manufacture of cables	11,561	11,327	11,603	12,404	12,384
Construction of telecommunications facilities	6,029	5,990	5,388	5,458	4,620
Software	2,457	3,162	4,014	5,488	6,913 *
Information software	1,190	1,402	1,817	2,254	2,767
Services incidental to communications	4,870	5,114	6,005	6,651	6,696

* Source: A Future Vision of Advanced Telecommunications Services," prepared by the Telecommunications Council and published by the MPT in 1987. These figures should be seen as representing the information services industry as a whole. The estimated figure for Type II carrier business was considerably less than 50 billon yen. Other source consulted are the Report on Specific Service Industries," published by the MITI and the Information Services Industry While Paper," prepared by the Information Services Industry Association.

** Both figures include the manufacture and lease of computers.

Appendix 1. Comparison of Liberalization Backgrounds

Country Item	Japan	Britain	U.S.A.
Revitalization of Industry All Sectors	O	O	O
Telecommunications	O	O	O
Administrative/ Financial Reform	O	△*	X
Privatization	O	O	X
Deregulation in General	△	△	O
Antitrust Law Problem	X	X	O
Competition Preceded Deregulation?	X	X	O
Key Words	Privatization Introduction of Principles of Competition	Privatization Liberalization	Deregulation Divestiture

O: yes
△: to limited extent
X: no

* Rationalization of Nationalized Industry

Appendix 2. Comparison of Legislative Environments

Country / Criteria	Japan	Britain	U.S.A.
Scope of Liberlization			
Local	O	O	X *
Long Distance	O	O	O
International	O	O	O
Paging & Others	O	O	O
Basic Concept on which Regulation depends	Business	Service	Service
Classification of Business or Service	Type I Type II	Basic VAN	Basic Enhanced
Scope and Degree of Regulation	Type I most strict General Type II most relaxed	Basic strict VAN relaxed	Only Basic submitted to Regulation Enhanced not Regulated
Information Processing industry Included or Excluded	Included to the extent it coincides with Type II	Included to the extent it coincides with VAN	Excluded
Number of Carriers	Fairly large both in Type I and Type II	Basic...only 2 at present and in the near future	Fairly large

Country / Criteria	Japan	Britain	U.S.A.
Treatment for the Dominant Carrier	NTT has special obligations for universal service and R&D	BT has the universal services obligation.	No universal service obligation, but asymmetric regulation between AT&T and OCCs.
Principles serving as Base for Tariff Authorizatio	Principles described in Business Law Totally depends upon ROR principle	RPI-3% rule with some exceptions	ROR principle
Ratio of the Tariff for Local Calling to the Most Expensive Tariff for Toll Calling	1 : 36	1 : 4	1 : 16
Regulatory Bodies	Ministry of Posts & Telecommunications (MPT)	Ministry of Trade and Industry Office of Telecommunications (OFTEL)	Federal Communications Commission (FCC) Judge Greene (Court of Appeal) Department of Justice (DOJ)
Related Agencies	Ministry of Int'l Trade and Industry (MITI), as far as it relates to Information Processing Industry	-------	National Telecommunications and Information Agency (NTIA) State Public Utility Commissions *

* Intra LATA service is regarded as one of the public utilities and under the jurisdication of local PUCs.

References

1. Crozier, Michel.

   Experiments in Administrative Modernization in the U.S., Sweden and Japan. To appear in October 1988.

2. Hayashi, Koichiro.

   Liberalization of the Telecommunications Industry in Japan. Presented at the EARIE Conference, August 1986.

3. ________________.

   The Economies of Networking - Implications for Telecommunications Liberalization. First presented at the ITS Conference, June 1988. Second presented at the IIC Conference, September 1988.

4. Kojima, Masashi.

   An Assessment of the Changes in Japanese Telecommunications Business. Presented at the World Telecommunications Symposium, December 1986.

5. NTT.

   Telecommunications Reform and NTT Privatization. NTT Publication G115. February 1985.

REPEATED AUCTIONS OF INCENTIVE CONTRACTS, INVESTMENT, AND BIDDING PARITY WITH AN APPLICATION TO TAKEOVERS*

by

J.-J. Laffont**

and

J. Tirole***

October 1988

---

* Support from Commissariat du Plan, the Pew Charitable Trust, and the Center for Energy Policy Research at MIT is gratefully acknowledged. We thank Bernard Caillaud, Oliver Hart, Benjamin Hermalin and James Poterba as well as the editor, an associate editor and a referee.

** GREMAQ and California Institute of Technology

*** M.I.T.

REPEATED AUCTIONS OF INCENTIVE CONTRACTS, INVESTMENT, AND BIDDING PARITY WITH AN APPLICATION TO TAKEOVERS

J.-J. Laffont and J. Tirole

ABSTRACT

This article considers a two-period model of natural monopoly and second sourcing. The incumbent supplier invests in the first period. After observing the incumbent's first-period performance, the buyer may breakout in the second. The investment may or may not be transferable to the second source, and it may be monetary or take the form of human capital.

We determine whether the incumbent should be favored at the reprocurement stage, and how the slope of his incentive scheme should evolve over time. It emerges from our analysis that the gains from second sourcing are not as high as one might hope.

Finally, we reinterpret the second source as a raider and the breakout as a takeover. We discuss the desirability of defensive tactics and obtain a rich set of testable implications concerning the size of managerial stock options, the extent of defensive tactics, the firm's performance, and the probability of a takeover.

JEL: 026, 512, 619

Keywords: Incentives, Second Sourcing, Bidding Parity, Auctions, Regulation

REPEATED AUCTIONS OF INCENTIVE CONTRACTS, INVESTMENT, AND BIDDING PARITY WITH AN APPLICATION TO TAKEOVERS

J.-J. Laffont and J. Tirole

I. INTRODUCTION

The regulation of a natural monopoly is often a repeated game. If the regulated monopoly's performance is not adequate, it may be in the regulator's interest to look for another firm (or team of managers) to replace the incumbent. Second sourcing indeed occurs in the reprocurement of defense contracts, in the repeated bidding for franchises, and in private contracting. Should auctions be set up that sequentially select the regulated firm? Should such auctions be concerned with bidding parity between the incumbent and the entrants? What is the incumbent firm's incentive to invest in physical capital? in human capital? The determination of the optimal breakout policy and its interaction with incentive schemes are the topics of this article.

The "Chicago approach" to regulating a natural monopoly (Demsetz (1968), Stigler (1968), Posner (1972)) suggests that a monopoly franchise be awarded to the firm that offers to supply the product on the best terms. Franchise bidding may also be repeated over time to adjust for new, non-contracted-for circumstances or to encourage entry of another, more efficient firm. It allows the regulator to select the most efficient supplier at any point in time. Williamson (1976),[1] responding to this approach, has forcefully made the following points:[2]

1) Physical capital and, even more, human capital are not always easily transferable from one firm to the other. Hence, symmetry

between the firms is unusual at the franchise renewal stage. The incumbent enjoys an advantage over its competitors.

2) Even when the incumbent's capital is transferable, the corresponding investment is hard to measure. This is clearly true for non-monetary investments; for instance, the quality of past investment choices admits no monetary measure. Furthermore, accounting manipulations garble the measurement of monetary investments. For instance, the incumbent can integrate into supply or arrange kickbacks from the equipment suppliers; or he can affect depreciation charges. The prospect of possibly being replaced by an entrant lowers the incumbent's incentive to invest in capital that he will not be able to transfer at the right price.[3]

These two points form the building blocks of our model. We assume that part of the incumbent's investment is transferable (general) and part is non-transferable (specific) (point 1). Furthermore the regulator can observe the regulated firm's cost or profit but is unable to recover the precise amount of investment from aggregate accounting data (point 2).

The model has two periods. In the first, the regulator offers an incentive contract to a single firm (the incumbent, the first source). The incumbent's cost, which is the only variable observed by the regulator, is a function of the firm's intrinsic productivity or efficiency, the firm's first period "effort", and a monetary investment. By a simple relabelling of variables, the model also allows for a non-monetary investment. The firm knows its productivity and chooses both effort and investment. In the second period (reprocurement stage), the regulator can keep the incumbent or invite another firm (the entrant, the second source) to replace the incumbent. The entrant's intrinsic productivity is known to the entrant only and can be higher or lower than the incumbent's. The second-period cost of the selected firm depends on its efficiency, its second-period effort, and the first-period investment (the entire investment for the incumbent but only the transferable part for the entrant). We assume that the incumbent's and the entrant's efficiency parameters are drawn from a common distribution; any ex ante intrinsic discrepancy in second-period efficiency is attributable only to non-transferable investment.

We focus on the interaction between the breakout rule and the intertemporal evolution of the slope of the incumbent's incentive scheme. (Optimal incentive schemes turn out to be linear in cost.) We say that *bidding parity* obtains when the regulator selects the entrant if and only if his *second-period* efficiency exceeds the incumbent's, where these efficiencies include the effect of the first-period investment. The regulator favors the incumbent when he selects the entrant only if the entrant is sufficiently more efficient than the incumbent, and vice versa. The Chicago school's recommendation alluded to above is tantamount to advocating bidding parity.

In the case of a transferable investment, any cost savings from the investment become savings for the second source if the first source is replaced. Because the investment is not observable and so cannot be compensated directly, the incumbent has too little incentive to invest; with some positive probability, the fruits of this costly investment will accrue to the second source. The regulator optimally uses both the breakout rule and the intertemporal evolution of the incentive scheme to correct this externality. First, he favors the incumbent at the reprocurement stage to increase his incentive to invest. To see why this is optimal, note that a small departure from bidding parity implies only a negligible loss in ex-post productive efficiency. Second, he provides for time-increasing incentives. The incumbent bears a small fraction of his first-period cost and therefore perceives investment as cheap. In contrast, he bears a high fraction of his second-period cost (if his contract is renewed), which ex ante gives him much incentive to invest.

The policy recommendations are quite different in the case of a *non-transferable* investment. In this case, the incumbent fully internalizes the social value of his cost savings, so that the previous externality does not exist. There is, however, a new effect: the incumbent has "on average" a cost advantage at the reprocurement stage. Our result, reminiscent of the theory of auctions with asymmetric bidders, is that if the distribution of potential efficiencies satisfies the classic monotone hazard rate property, the auction should be biased in favor of the higher cost firm--the entrant. The slope of the incumbent's incentive scheme should be time-invariant.

In Section II we set up the model and in Section III solve for the optimal breakout rule and incentive schemes. Next we study a variant of the model in which the first-period effort to reduce first-period cost also reduces the second-period cost. This "learning-by-doing" investment does not increase the first-period cost, in contrast with the basic model. It leads the regulator to favor the incumbent at the reprocurement stage (Section IV). Section V summarizes the results obtained thus far and assesses their relevance. In Section VI we focus on transferable investment and obtain a rich set of policy recommendations and testable implications. We also show how the optimal regulatory policy can be implemented through the set-up of a second-period auction combined with a simple incentive package composed of "performance related incentives" (cost reimbursement rules) and "switching incentives" (cancellation fee or entry fee). For expositional simplicity, we relegate most proofs of Sections III through VI to the Appendix and give only the intuition in the text.

Our model of second sourcing can be reinterpreted in terms of takeovers and managerial myopia, and this we do in Section VII. The incumbent firm becomes the incumbent management team; the second source is the raider (rival management team). Cost is reinterpreted as profit. Favoring the incumbent at the contract renewal stage corresponds to allowing certain defensive tactics. Our pessimistic assessment of second sourcing translates into a qualification of the economists' recent partial view of takeovers as a managerial discipline device. We also show that in an optimal managerial contract, the firm's performance, the manager's golden parachute, and the manager's stock options are positively correlated. These three variables are negatively correlated with the probability of a takeover. Concluding remarks appear in Section VIII.

Before proceeding we would like to acknowledge the earlier literature on intertemporal procurement of a single firm under commitment (Baron-Besanko (1984)), second sourcing (Anton-Yao (1987), Caillaud (1985), Demski et al (1987), Scharfstein (1988)), and auctions of incentive contracts (Laffont-Tirole (1987b), McAfee-McMillan (1987), Riordan-Sappington (1987)). Although none of these papers considers simultaneously intertemporal regulation, investment, and second sourcing, and therefore is apt to address the issue of bidding parity, we make considerable use of their insights.

II. THE BASIC MODEL

We consider a two-period natural monopoly model. Each period a project valued S by consumers must be realized by a single firm, but the identity of the firm can change over time. In period 1, there is a single firm, the incumbent, with cost function:

$$C_1 = \beta - e_1 + d(i). \tag{1}$$

Here $\beta \in [\underline{\beta}, \bar{\beta}]$ is the incumbent's intrinsic cost parameter, e_1 is the level of effort exerted by the firm's manager, and $d(i)$ $(d' > 0,\ d'' > 0)$ is the cost of a first-period (monetary) investment that lowers the incumbent's second period cost by i. The disutility of effort for the incumbent is $\psi(e_1)$ with $\psi' > 0$,

$\psi''>0$, $\psi'''\geq 0$.[4] The incumbent's second-period cost function is:

$$C_2=\beta-e_2-i \tag{2}$$

where β is the same parameter as in period 1 and e_2 is the effort exerted in period 2, at disutility $\psi(e_2)$.

The investment $d(i)$ in period 1 also lowers the entrant's (second period) cost by ki, where $k\in[0,1]$. If $k=0$, the investment is *non-transferable* (or specific). If $k=1$, the investment is *transferable* (or general). More generally, we allow the investment to be partly transferable and partly specific. Hence, the (potential) entrant has in period 2 a cost function:

$$C'=\beta'-e'-ki, \tag{3}$$

where $\beta'\in[\underline{\beta},\bar{\beta}]$ is the entrant's intrinsic cost parameter, which is learned at the beginning of period 2, and e' is his level of effort. The entrant has the same function of disutility of effort as the incumbent. An additive uncertainty could be added to the costs without any change in our results, as will be clear from the linearity in cost of optimal contracts. All costs are accounting costs, which are commonly observed. The parameters β and β' are independently drawn from the same distribution with cumulative distribution function $F(\beta)$ and density function $f(\beta)$ continuous and positive on $[\underline{\beta},\bar{\beta}]$ with $\frac{d}{d\beta}[F(\beta)/f(\beta)]\geq 0$.[5]

Remark on non-monetary investments: In the introduction, we claimed that our model depicts both monetary and non-monetary investments. The formulation above assumes a monetary investment, as $d(i)$ increases the first-period cost. To substantiate our claim, suppose that there is no monetary investment. The first-period cost is then

$$C_1=\beta-\tilde{e},$$

where $\tilde{e}$ is the effort expended to reduce the first-period cost. The incumbent's managers exert a second type of effort that reduces the incumbent's second-period cost by $\tilde{\tilde{e}}$ and the entrant's cost by $k\tilde{\tilde{e}}$. The first-period effort to obtain this cost reduction is $d(\tilde{\tilde{e}})$, so that the total first-period disutility of effort is $\psi(\tilde{e}+d(\tilde{\tilde{e}}))$. One may think of $d(\tilde{\tilde{e}})$ as the number of hours spent in the first period on finding the right technology for the second period. A simple change in variables shows that this model is formally equivalent to the monetary investment model set up above: let $e_1\equiv\tilde{e}+d(\tilde{\tilde{e}})$ denote the total first-period effort and $i\equiv\tilde{\tilde{e}}$. Then, the first-period cost can be written:

$$C_1=\beta-(e_1-d(e))=\beta-e_1+d(i).$$

The regulator's problem is to organize production and transfers so as to maximize social welfare. The incumbent's expected utility level is:

$$U=t-\psi(e_1)-\delta\pi\psi(e_2) \qquad (4)$$

where δ is the discount factor, π is the probability that the incumbent will remain active in period 2, and $t\equiv t_1+\delta t_2$ is the net expected (present discounted) transfer received by the firm from the regulator. The firm receives t in addition to the payment of realized costs by the regulator. We will assume that the regulator has the ability to commit to two-period regulatory schemes.[6] To obtain the incumbent's participation the regulator must ensure that[7]

$$U\geq 0, \qquad (5)$$

where the individual rationality (IR) level has been normalized to zero.

The entrant's utility level, if the entrant is active in period 2, is:

$$V = t' - \psi(e') \tag{6}$$

where t' is the (expected) net transfer received from the regulator. The entrant's I.R. constraint is:

$$V \geq 0. \tag{7}$$

A utilitarian regulator wishes to maximize the sum of expected utilities of consumers and firms. Under complete information his instruments are the levels of investment, efforts, transfers, and the *breakout* rule (i.e., the choice of who is active in period 2). Complete information implies that the (potential) entrant should be allowed to enter if and only if $\beta' < \beta^*(\beta, i) = \beta - (1-k)i$.

Letting λ denote the shadow cost of public funds, the consumers' expected utility level is:

$$\begin{aligned} S - (1+\lambda)(C_1 + t_1) + \delta(1 - F(\beta^*(\beta, i)))[S - (1+\lambda)(C_2 + t_2)] \\ + \delta\int_{\underline{\beta}}^{\beta^*(\beta, i)} (S - (1+\lambda)(C'(\beta') + t'(\beta')))]f(\beta')d\beta', \end{aligned} \tag{8}$$

where we assume that consumers have the same discount factor as the incumbent.

Since $\lambda > 0$, the IR constraints (5) and (7) are binding in the regulator's optimization program. The regulator's program thus reduces to:[8]

$$\underset{\{e_1, e_2, e', i\}}{\text{Max}} \{S(1+\delta) - (1+\lambda)(\beta - e_1 + d(i) + \psi(e_1)) \tag{9}$$

where e_1, e_2, and i are functions of β, and e' is a function of β'. This is a quasi-concave problem[9] with first-order conditions:

$$\psi'(e_1) = \psi'(e_2) = \psi'(e') = 1 \tag{10}$$

$$d'(i) = \delta[(1 - F(\beta - (1-k)i)) + kF(\beta - (1-k)i)]. \tag{11}$$

Equation (11) tells us that investment should be set at the level that equates its marginal cost, $d'(i)$, with its expected social marginal utility which is δ if the investment is transferable (k=1), but only $\delta(1-F(\beta-i))$ if it is not (k=0). The *positive externality* of the first-period investment on the entrant's cost must be internalized when it exists. Equation (10) indicates that the marginal disutility of each type of effort must be equated to its marginal benefit.

In the next section we solve the regulator's optimization problem under incomplete information.

III. OPTIMAL REGULATION UNDER ASYMMETRIC INFORMATION

Suppose that the regulator observes costs but does not know the parameters β and β' (even though it knows their distribution) and cannot observe effort levels. We also assume that the investment is unobservable by the regulator. (The case of observable investment will be treated as a special case.)

The regulator's instruments are now cost reimbursement rules as well as a breakout rule. Using the revelation principle, this regulatory procedure can be viewed as a revelation mechanism for the incumbent composed of a contract $(C_1(\beta), C_2(\beta), t(\beta))$ and a breakout rule $\beta^*(\beta)$ and as a revelation mechanism for the entrant $(C'(\beta'/\beta), t'(\beta'/\beta))$. In this section, we use the abstract framework provided by the revelation principle to characterize the optimal allocation (efforts, investment, breakout rule). In Section VI, we show how this optimal allocation can be implemented through familiar mechanisms like linear cost reimbursement rules, auctions, and cancellation fees, and we develop some further empirical implications of the optimal regulatory scheme.

Before proceeding we should make a methodological point concerning the amount of information received by the entrant about the incumbent's productivity. The revelation principle tells us that, without loss of generality, the principal may ask the incumbent truthfully to announce his type: $\tilde{\beta}=\beta$. Should $\tilde{\beta}$ be revealed to the entrant? If it is not, the incumbent's first-period cost still reveals information about the incumbent's productivity. Is it worth distorting the first-period allocation to garble the entrant's information about β? For instance, if the optimal first-period regulation implies that C_1 perfectly reveals β (as will be the case here), would one want to induce some first-period pooling so that the entrant would possess less information about β than the principal and possibly would bid more aggressively? Fortunately the answer is no. Maskin and Tirole (1986), in their study of contracts designed by an informed principal (here the regulator), show that if preferences are *quasi-linear* (as is the case in this paper), the design of the contract for the entrant does not depend on whether or not the agent (here the entrant) knows the principal's information. Hence there is no point hiding the announcement $\tilde{\beta}$ from the entrant or distorting the first-period allocation.[10]

Let $U(\beta)$ be the rent extracted by an incumbent of type β. Following Laffont and Tirole (1987a), the incumbent's incentive compatibility and individual rationality constraints can be written:[11]

$$\dot{U}(\beta)=-\psi'(\beta-C_1(\beta)+d(i))-\delta(1-F(\beta^*(\beta)))\psi'(\beta-C_2(\beta)-i) \tag{12}$$

$$U(\bar{\beta})=0 \tag{13}$$

with sufficient second-order conditions[12]

$$\frac{dC_1}{d\beta} \geq 0; \frac{dC_2}{d\beta} \geq 0; \frac{d\beta^*}{d\beta} \geq 0 \, . \tag{14}$$

Equation (12) results from the fact that when the incumbent's intrinsic cost parameter is reduced by 1, the incumbent can reduce his effort by 1 in each period and obtain the same transfers. His rent is thus increased by $[\psi'(e_1)+\delta\pi\psi'(e_2)]$ where $\pi=(1-F(\beta^*(\beta)))$ is the probability of no breakout. Equation (13) states that the least efficient type gets no rent at the optimum.

Moreover, since investment is non-observable we must consider the (moral hazard) constraint[13] describing the incumbent's choice of investment:

$$-d'(i)\psi'(\beta - C_1(\beta) + d(i)) + \delta(1 - F(\beta^*(\beta)))\psi'(\beta - i - C_2(\beta)) = 0. \tag{15}$$

Suppose that the firm contemplates a unit increase in investment. The extra investment cost is $d'(i)$. To keep the first-period cost (and thus reward) constant, effort must be increased by the same amount, thereby engendering additional disutility $\psi'(e_1)d'(i)$. In turn, the second-period effort can be reduced by 1, saving disutility $\psi'(e_2)$, with probability $1-F(\beta^*(\beta))$.

Similarly, let $V(\beta'/\beta)$ be the rent extracted by an entrant of cost characteristic β' when the incumbent's type is β. The entrant's incentive and IR constraints can be written:

$$\frac{\partial V}{\partial \beta'}(\beta'/\beta) = -\psi'(\beta' - ki - C'(\beta'/\beta)) \tag{16}$$

$$V(\beta^*(\beta)/\beta) = 0 \, , \tag{17}$$

with the second order condition $\frac{dC'}{d\beta'} \geq 0$. Integrating (16) with (17) as a boundary condition yields

$$V(\beta'/\beta) = \int_{\beta'}^{\beta^*(\beta)} \psi'(\tilde{\beta} - ki - C'(\tilde{\beta}/\beta))d\tilde{\beta}. \tag{18}$$

Using U and V as state variables, the regulator's maximization problem can be written as the following optimal control problem:

$$\begin{aligned}\text{Max}\int_{\underline{\beta}}^{\beta} \{ & S(1+\delta)-(1+\lambda)(C_1(\beta)+\psi(\beta-C_1(\beta)\ +d(i)))-\lambda U(\beta) \\ & -\delta(1-F(\beta^*(\beta))(1+\lambda)(C_2(\beta)+\psi(\beta-i-C_2(\beta))) \\ & -\delta\int_{\underline{\beta}}^{\beta^*(\beta)}[(1+\lambda)(C'(\beta'/\beta)+\psi'(\beta'-ki-C'(\beta'/\beta))+\lambda V(\beta'/\beta))f(\beta')d\beta'\}f(\beta)d\beta\end{aligned} \tag{19}$$

subject to: (12)(13)(15)(16)(17).

In this formulation, the regulator's objective function has been rewritten to separate the efficiency costs of the form $(1+\lambda)(C+\psi)$ and the "distributional" costs λU and λV.

Remark 1: Since some second-order conditions are only sufficient, we do not impose them but rather check later that they are satisfied by our solution.

Remark 2: We do not allow the incumbent's contract to depend on the entrant's realized cost C' following a breakout. In the case of non-transferable investment, this does not involve any loss of generality because C' does not contain any information about i. It does involve a loss of generality, however, for transferable investments (an artifact of the risk neutrality framework is that such a dependence would enable the first best level of investment to be achieved with sufficiently large penalties). Excluding such a dependence of t on C' is a good approximation to reality for several reasons. First, this dependence would create a delayed transfer or penalty. The displaced incumbent would be required, for instance, to pay a penalty 5 or 10 years after the breakout, which raises the issue of the feasibility of such long-run contracts; in contrast, our transfers can follow production immediately. Second, and maybe more importantly, the entrant's cost may be subject to manipulation. Indeed, ex post, the entrant and the regulator have an incentive to tinker with accounting data on C' so as to force the incumbent to pay a penalty. Hence, letting the incumbent's reward depend on the entrant's cost may not be feasible after all. Third, in the presence of cost uncertainty and risk aversion (or, alternatively, limited penalties), the incumbent would not bear a large fraction of its successor's realized cost. To summarize, although some dependence of the incumbent's reward on post-termination performance may be feasible in some instances, it is at most a limited instrument, which leaves ample scope for the policies described in this article.[14]

Solving the problem in (19) we obtain:

Proposition 1: With transferable and unobservable investment, the optimal breakout rule favors the incumbent ($\beta^*(\beta)<\beta$) *except at* $\beta=\underline{\beta}$, *where bidding parity obtains:* $\beta^*(\underline{\beta})=\underline{\beta}$. *Bidding parity* ($\beta^*(\beta)=\beta$) *holds when the investment is transferable and observable.*

Proof: See Appendix.

The intuition behind this result is straightforward. When investment is transferable and observable, the regulator can directly force the incumbent to internalize the positive externality on the entrant's cost by imposing the right level of investment. Since the firms' efficiencies are drawn from the same distributions, bidding parity should hold. Incentive questions remain separate from the breakout rule. If investment is unobservable, there is underinvestment because the incumbent has no reason to internalize the externality. The regulator then mitigates this inefficiency by raising the probability that the incumbent remains active in period 2. This is achieved by favoring him in the breakout rule. An incumbent of type β has a zero probability of being replaced. He thus invests the socially optimal amount, and the selection rule need not be biased to encourage investment.

The next result is that the regulator alters the incentive schemes as well as the breakout rule to remedy this externality problem. This is best explained by replacing the non-linear transfer functions $t(C)$ by linear contracts as in Laffont and Tirole (1986).[15] The transfers can then be written (see proof of Proposition 2)

$$t_1(C_1^a, C_1) = G_1(C_1^a) - K_1(C_1^a)(C_1 - C_1^a)$$

for the incumbent in period 1,

$$t_2(C_2^a, C_2/C_1^a) = G_2(C_2^a, C_1^a) - K_2(C_2^a)(C_2 - C_2^a)$$

for the incumbent in period 2 (in the case of no breakout), and

$$t'(C'^a, C'/C_1^a) = G'(C'^a, C_1^a) - K'(C'^a)(C' - C'^a)$$

for the entrant in period 2 (in the case of breakout), where "a" denotes an expected value; t_1, t_2, and t' are the net transfers (after cost reimbursement); G_1, G_2, and G' are the fixed components of the transfers; K_1, K_2, and K' are the slopes of the incentive schemes with respect to the realized costs, which are equal to 0 for cost-plus contracts and to 1 for fixed-price contracts.

The interesting questions refer to the slopes of the incentive schemes. We obtain

Proposition 2: With unobservable and at least partly transferable investment ($k>0$), the incumbent's first-period incentive scheme is "low-powered" relative to the second-period one ($K_1<K_2$). Because the slope with observable investment is time-invariant and lies between K_1 and K_2, unobservability

calls for a flatter incentive scheme in the first period and a steeper incentive scheme in the second period than does the observable-investment scheme.

Proof: See Appendix.

The intuition is that a contract resembling more a cost-plus contract in the first period lowers the incumbent's investment cost, while a contract closer to a fixed price contract in the second period allows the incumbent to capture the proceeds of the investment.

The results are quite different when investment is specific.

Proposition 3: (i) With unobservable non-transferable investment, the optimal breakout rule favors the entrant $(\hat{\beta}^*(e)>\beta-i)$. *(ii) The unobservability of non-transferable investment imposes no cost on the regulator.*

Proof: See Appendix.

Part (ii) of Proposition 3 is unsurprising. A classical result in incentive theory is that the slope of the incumbent's incentive scheme must be time-invariant if investment is observable (see the proof of Proposition 3). Because there is no externality between the incumbent and the entrant when investment is non-transferable, and because the incumbent's and the regulator's intertemporal preferences coincide (due to the constant slope and identical discount factors), the incumbent correctly selects the socially optimal investment; that is, unobservability of investment imposes no social loss.

Part (i) of Proposition 3 is more subtle than part (ii) and relies on the monotone hazard rate assumption. To understand it, recall the static single firm regulatory problem. Let β be drawn from a cumulative distribution function $F(\cdot)$, with density $f(\cdot)$ on $[\underline{\beta},\bar{\beta}]$. When choosing an incentive scheme for the firm, the regulator must trade off efficiency, which would call for a fixed price contract, and minimization of the firm's informational rent, which would call for a cost-plus contract. This tradeoff yields a distortion in the effort allocation for all $\beta>\underline{\beta}$. By reducing the distortion in effort for parameter β, the regulator realizes a gain proportional to $f(\beta)$. At the same time, it must give a higher rent to all types of firms that are more efficient than β, in proportion $F(\beta)$, because the latter can always mimic the behavior of a less efficient firm. At the optimum, the marginal gain in efficiency must equal the marginal cost associated with the firm's expected rent. Now consider our two-period model. Suppose that the incumbent's and the entrant's productivity parameters β and β' are independently drawn from the same distribution $F(\cdot)$, that is, any observable discrepancy in intrinsic efficiency is attributable to the incumbency advantage. Suppose that, in the second period, the regulator does not favor the incumbent or the entrant, and consider parameters β and β' such that the two firms have the same *second-period* intrinsic efficiency. In the presence of specific investment, the "cut-off point" is $\beta'=\beta^*(\beta)=\beta-i<\beta$. Thus, if we make the classic assumption that the hazard rate F/f is an increasing function, one has $F(\beta^*)/f(\beta^*)<F(\beta)/f(\beta)$. This means that at equal second-period intrinsic efficiency, the optimal regulation of the entrant calls for less distortion of effort than does regulation of the incumbent. Another way of phrasing this intuition begins by noticing that the selection of a firm amounts to an upward truncation of the distribution of its productivity parameter. Thus, at equal second-period intrinsic efficiency, the regulator is less uncertain about the entrant's productivity than about

the incumbent's and therefore can regulate the entrant more efficiently. Specific investments thus call for favoring the entrant at the reprocurement stage. We call this effect the *rent differential effect*.

An interesting analogy with the literature can be drawn here. Demski, Sappington, and Spiller (1987) offer a second-sourcing example in which the purchaser selects the entrant rather than the incumbent to be the producer, even though he knows that the incumbent has lower production costs. (Corollary 5, page 91. For similar results, see Caillaud [1985] and, in the classic context of auctioning of an object, Myerson [1981] and McAfee-McMillan [1984].) The Demski et al. model does not allow any investment. But in this example the incumbent's and the entrant's production costs are drawn from asymmetric distributions, and the incumbent's cost distribution stochastically dominates the entrant's. Our model presumes identical cost distributions ex-ante, but the existence of specific investment confers a (statistical) superiority on the incumbent ex post. Like Demski et al., we find that this stochastic dominance by the incumbent calls for favoring the entrant.

Two last points conclude this section. First, in spite of the incentives to invest described in Propositions 1 and 2, it is still the case that the incumbent *underinvests relative to the social optimum*.[16] Second, a reasonable conjecture is that the bias in the breakout rule $(\beta-i)-(\beta^*(\beta)-ki)$ is an increasing function of k. We have proved this result only in the case of observable investment and k in a neighborhood of 1. We suspect the property holds more generally.

IV. LEARNING BY DOING

An interesting variant of our model allows learning by doing rather than investment to reduce the second-period cost. We now assume that the incumbent's effort in period 1, e_1, affects costs in period 2 so that

$$C_2 = \beta - e_2 - (a+b)e_1 \tag{20}$$

and

$$C' = \beta' - e' - ae_1. \tag{21}$$

where $a, b \geq 0$. That is, the fraction $\bar{k} \equiv a/(a+b)$ of such learning is transferable to the entrant. One may have in mind that the incumbent exerts effort that yields tricks or technologies that reduce his first-period production cost and that will still be used in period 2.

Under complete information the breakout rule is $\beta^*(\beta) = \beta - be_1$. The optimal effort levels would be determined by the program:

$$\begin{aligned} \underset{\{e_1, e_2, e'\}}{\text{Max}} \ &\{S(1+\delta) - (1+\lambda)(\beta - e_1 + \psi(e_1)) \\ &- \delta(1 - F(\beta - be_1))(1+\lambda)(\beta - e_2 - (a+b)e_1 + \psi(e_2)) \\ &- \delta(1+\lambda)\int_{\underline{\beta}}^{\beta - be_1}(\beta' - e' - ae_1 + \psi(e'))f(\beta')d\beta'\}. \end{aligned} \tag{22}$$

The first order conditions are:[17]

$$\psi'(e_1) = 1 + \delta a + \delta b(1 - F(\beta - be_1)) \tag{23}$$

$$\psi'(e_2) = 1 \tag{24}$$

$$\psi'(e') = 1. \tag{25}$$

In particular, equation (23) equates the marginal disutility of effort to its total marginal social benefit, which is the first period benefit, 1, plus the transferable effect on the second period, δa, plus the expected specific effect $\delta b(1 - F(\beta - be_1))$.

Under incomplete information we must add the incentive constraints in the regulator's optimization program. Following the same lines of argument as in Section IV we obtain:

Proposition 4: With unobservable and fully transferable learning by doing ($b=0$), the breakout rule favors the incumbent ($\beta^*(\beta)<\beta$).

Proof: See Appendix.

There are now two effects favoring the incumbent: the externality effect of Proposition 1 and the *learning by doing effect*. Recall that the incumbent's informational rent comes from the possibility of mimicking a less efficient type's cost by exerting less effort. Under learning by doing, a reduction in the first-period effort reduces the second-period efficiency and rent, and this all the more if the probability of keeping the franchise is high. Hence, by increasing the probability of choosing the incumbent in the second period, the regulator makes it more costly for the incumbent to hide his efficiency in the first period. This effect, which does not exist for monetary investment, calls for favoring the incumbent.

When learning is partly specific, the relevant comparison for $\beta^*(\beta)$ is with $\beta - be_1$. Now at $\beta^* = \beta - be_1$ the rent obtained by the incumbent is higher at the parity point, and this calls for favoring the entrant as in Proposition 3. It can be shown that either effect may dominate; see Laffont-Tirole (1987a). In particular, when the learning is purely specific ($a=0$), we have two opposite incentive effects, the learning by doing effect and the rent differential effect.

A striking feature of the analysis in Section III is that the slope of the incumbent's incentive scheme must be time-increasing to encourage him to invest. In contrast, when investment is embodied in the first-period effort, lowering the incumbent's cost of investing is achieved through a high, rather than low, slope of the first-period incentive scheme. That is, choosing a low first-period effort is costly to an incumbent who bears a high fraction of his first-period cost. Indeed, $K_1(\beta)$ exceeds 1 for efficient βs,while in the solution in Section III, $K_1(\beta)\leq 1$ for all β. The next proposition specializes the model to the uniform-quadratic case: β is distributed uniformly on $[\underline{\beta},\bar{\beta}]$ and $\psi(e)=e^2/2$ (the result is ambiguous for the more general model):

Proposition 5: With unobservable learning by doing, in the uniform-quadratic case, the first-period incentive scheme is steeper than the second-period one: $K_1(\beta)>K_2(\beta)$ for all β.

V. DISCUSSION

Table 1 summarizes the analysis of the optimal allocation:

TABLE 1
CHARACTERISTICS OF THE OPTIMAL ALLOCATION

Nature of Investment		Breakout Rule	Time-Invariance of Slope of Incentive Schemes
Transferable	Observable Investment	Bidding Parity	Yes
	Unobservable Monetary or Nonmonetary Investment	Incumbent Favored (Externality effect)	No ($K_1 < K_2$)
	Learning by Doing	Incumbent Favored (Externality effect + learning effect)	No
Non-transferable	Observable Investment	Entrant Favored (Rent differential effect)	Yes
	Unobservable Monetary or Nonmonetary Investment		
	Learning by Doing	Ambiguous (Two opposite effects: learning and rent diffential)	No

The previous analysis aimed at generality and allowed many technological patterns. Before proceeding, it is worth assessing the relevance of the various effects leading the regulator to bias the bidding process. It should be clear that our focus is on sizeable, nonobservable investment. When investments matter, breakouts are most likely to be observed (and to be socially desirable) when the entrant is not at too much of a cost disadvantage, i.e., when it benefits from the incumbent's investment. The incumbent should be favored at the reprocurement stage, however, precisely when investment is transferable. These two facts lead us to a somewhat pessimistic assessment of how much regulators can hope to gain by using second sourcing in a natural monopoly situation involving substantial investments. Although substantial intrinsic cost differences between the firms may justify breakouts, our recommendations are more likely to apply to industries with transferable investments. In the rest of the paper, we focus on such industries. Section VI derives some further recommendations and testable implications of our model while Section VII shows how our results apply to takeovers where, we would argue, most of the firm's assets are transferable to a second source, i.e., to a raider.

VI. TRANSFERABLE INVESTMENT

We now specialize the model of Section III to transferable investment, a quadratic disutility of effort, and a uniform distribution of cost parameters:

Assumption A:

A1) $k=1$

A2) $\psi(e)=e^2/2$

A3) $F(\cdot)$ is uniform on $[0,1]$

A4) $1>\lambda$.

Assumption A4 is technical and is consistent with the second-order conditions, which require that λ not be too large.

We can now state:

Proposition 6: Under Assumption A,

(i) *The bias in favor of the incumbent is higher, the less efficient the incumbent is; that is;*

$$\frac{d}{d\beta}(\beta-\beta^*(\beta))>0\ .$$

(ii) *The incumbent exerts more effort in the second period than the type of entrant that makes entry socially indifferent; that is,* $e_2(\beta)>e'(\beta^*(\beta))$.

Proof of Proposition 6: See the Appendix.

Part (i) of Proposition 6 conveys an important intuition. An inefficient incumbent is replaced with high probability and therefore invests little. The bidding process must then be biased considerably so as to encourage him to invest. Part (ii) compares $e_2(\beta)$ and $e'(\beta^*(\beta))$. We knew that $e_2(\beta)>e'(\beta)$. There is a second effect, however, coming from the fact that $\beta^*(\beta)<\beta$ and that e' is decreasing. The two effects work in opposite directions, but the first dominates.

We saw in Section III that transferable investments call for special incentives to invest: bias in the breakout rule and time-increasing incentives. An inefficient incumbent imparts a particularly high externality because of a high probability of breakout. Proposition 6 showed that the breakout rule is distorted more for inefficient incumbents. Proposition 7 is its counterpart for the distortion in the slopes of the incentive scheme: An inefficient incumbent should face incentives that rise sharply over time.

Proposition 7: Under Assumption A, if λ is "not too large," then

$$\frac{d}{d\beta}(K_2(\beta)-K_1(\beta))>0.$$

Proof: Available from the authors upon request.

Proposition 6 enables us to obtain some interesting results on the second-period bidding process. As in Laffont-Tirole [1987b], one can view this bidding process as a first- or second-bid auction in which each bidder (here each firm) bids for the right to choose from a menu of monopoly linear incentive contracts. Unlike in our earlier auction paper, the menus of contracts differ between the two competitors because of the asymmetry of the problem. Let $U_2(\beta)$ denote the incumbent's second-period rent associated with the right to choose from his menu of contracts. Similarly, let $U'(\beta')$ denote the entrant's second-period rent. From incentive compatibility, we know that:

$$dU_2/d\beta=-\psi'(e_2(\beta))$$

and

$$dU'/d\beta'=-\psi'(e'(\beta')).$$

Hence, $U_2(\cdot)$ and $U'(\cdot)$ are defined up to positive constants. Although we will be mainly interested in their derivatives, we normalize these functions by imposing second-period individual rationality constraints,

$$U_2(\bar{\beta})=0$$

and

$$U'(\beta^*(\bar{\beta}))=0.$$

Recall that the highest β' who may be allowed to produce is $\beta^*(\bar{\beta})$.

Now consider a second-period first- or second-price auction in which each firm bids for the right to choose from its menu of linear incentive schemes. For simplicity, we treat the case of a second-price auction. From the usual equivalence theorem, we know that the first-price auction yields the same outcome. Then the incumbent bids $U_2(\beta)$ and the entrant bids $U'(\beta')$. In general, the equation $U_2(\beta)=U'(\beta')$ yields $\beta' \neq \beta^*(\beta)$ so that the second-price auction does not necessarily select the right firm. *The auction must thus be biased.* One way of doing so is to introduce a "golden parachute" or "cancellation fee" $G(\beta)$ to be paid to the incumbent if he is replaced. Hence, one can envision a first-period contracting process in which the incumbent chooses a first-period incentive scheme and a second-period golden parachute. The second-period allocation is determined by the above described auction.

For the right firm to be selected, the golden parachute must satisfy:

$$U_2(\beta)-G(\beta)=U'(\beta^*(\beta)),$$

as the incumbent shades his second-period bid by $G(\beta)$. Thus,

$$G(\beta)=\int_{\beta}^{\bar\beta}\psi'(e_2(x))dx-\int_{\beta^*(\beta)}^{\beta^*(\beta)}\psi'(e'(x))dx$$

or

$$G(\beta)=\int_{\beta}^{\beta}[\psi'(e_2(x))-\psi'(e'(\beta^*(x)))\frac{d\beta^*}{d\beta}(x)]dx.$$

We have:

Proposition 8: Under Assumption A, the golden parachute is positive, and it increases with the incumbent's efficiency (that is, $G(\beta)\geq 0$, $\dot{G}(\beta)<0$).

Proof of Proposition 8: For a quadratic disutility of effort, one has:

$$G(\beta) = \int_{\beta}^{\beta} \left(e_2 - e' \frac{d\beta^*}{d\beta}\right) dx$$

Proposition 6 implies that $e_2 > e' \frac{d\beta^*}{d\beta}$.

Q.E.D.

As we mentioned earlier, the important result in Proposition 8 is that the golden parachute increases with the firm's efficiency.

To summarize, the optimal allocation can be implemented by a second-period auction in which each firm bids for the right to be the monopoly supplier. Efficient selection is obtained by offering in the first-period a golden parachute together with a first-period incentive scheme. The golden parachute is characterized in Proposition 8.

Rather than using cancellation fees (golden parachutes), the regulator can have the entrant pay an entry fee to bias the auction. Let $P(\beta)$ denote this entry fee.[18] (P stands for "poison pill," because, as a rough description, a poison pill forces a raider to pay an extra-price to acquire the firm.) In the context of our model, an entry fee is equivalent to a negative cancellation fee, though in a more general model, the two instruments would not be perfect substitutes. Reinterpreting Proposition 8, we see that the entry fee should decrease with the incumbent's efficiency ($\dot{P}(\beta)>0$): An efficient incumbent should be protected relatively less than an inefficient one.

We summarize in Table 2 our recommendations and testable implications for transferable investments. The variables in class A are positively correlated among each other and negatively correlated with the variables in class B.

TABLE 2

SUMMARY OF IMPLICATIONS FOR TRANSFERABLE INVESTMENTS

A	B
• Incumbent's performance (minus cost, profit)	• Probability of second sourcing
• Slope of the incumbent's incentive schemes (K_1,K_2)	• Intertemporal increase in the slope of the incumbent's incentive $(K_2 - K_1)$
• Cancellation fee	• Entry fee

VII. A REINTERPRETATION: TAKEOVERS AND MANAGERIAL MYOPIA

As mentioned in the introduction, our model of second sourcing can shed some light on the desirability of takeovers. The entrant can be reinterpreted as a raider, the incumbent as the current managerial team. The accounting cost stands for per-period performance (profit or stock value). The cost parameter (β) is a measure of the inefficiency of current management, and the effort variable (e) refers to the possibility of self-dealing management (appropriation of profits, luxurious offices, personal jets, golf playing...). The reprocurement stage can be thought of as a tender offer.[19] The rigging of bidding parity in favor of the incumbent or the entrant is a rough formalization of defensive tactics and protakeover measures respectively.[20]

Our assumption that the incumbent's incentive scheme is not contingent on the entrant's performance translates into the assumption that the displaced managerial team does not keep substantial stock options in the firm after leaving. This assumption is made in most of the literature on the market for corporate control (e.g., Blair et al [1986], Grossman-Hart [1987], Harris-Raviv [1987]). Theoretical reasons can be found to motivate it. Although the arguments advanced in the context of regulation (in particular, the collusion argument in Section III) fare less well in this context if the managers are risk averse, the raider and displaced managers have an ex post incentive to renegotiate former contracts and may let the displaced managers resell stock options, which no longer serve an incentive purpose and create an excessive risk in the displaced managers' portfolio.[21]

Another effect limiting the role of post-termination stock options is that if there is no merger (so that the firm's stock remains traded), minority freeze-out problems are very acute. In this case, covenants must be imposed that greatly reduce the information value of the firm. (See Holmström-Tirole (1987) for an informal development of this point). At a more empirical level, this assumption also makes some sense. First, since many acquired firms do not have outstanding shares after the takeover, the imcumbent managers automatically exercise their stock options. Second, even if the raider acquires only control, many managerial contracts specify that the managers must exercise their options within 90 days if their employment is terminated. (Hence, in the context of our model, the options are exercised well before the investment pays off.) We believe that stock options encourage the incumbent managers to internalize the positive externality of *observable* investment on the raiders' post-takeover performance. Our point is that stock options provide very insufficient incentives to induce managers to internalize the effect of investments that are *not observable* by the market.[22]

Some implications of our model in the takeover context are (assuming that investment is transferred to the raider and is not of the learning-by-doing type):[23]

1. Firm performance and the probability of takeover are negatively correlated.
2. The use of defensive tactics to disadvantage the raider benefits the firm's shareholders.
3. The managers are given linear incentive schemes, which can be interpreted as stock options. Their incentive package also includes golden parachutes or poison pills.

4. The incumbent manager's stock options increase over time.
5. The size of the golden parachute is positively related to the number of stock options.
6. The size of the golden parachute is positively related to the firm's performance.[24]
7. The size of the poison pill is negatively correlated with the incumbent manager's stock options.
8. The size of the poison pill is negatively correlated with the firm's performance.

Conclusion 2 suggests that defensive tactics are not necessarily harmful.[25] Although most of the incentive literature on the topic views takeovers as a managerial discipline device, we believe that the popular fear of managerial myopia should not be neglected by economists. Our argument supplies an *efficiency* reason for foreclosing entry. That is, a social planner, whose objective function puts equal weight on the incumbent and the entrant biases the auctioning process against the entrant. When the principal is a private entity (as is the case for shareholders), the contract signed between the principal and the incumbent does not internalize its effect on the entrant's welfare. Aghion and Bolton [1987] have shown that the desire to extract the entrant's rent leads the two initial parties to sign a contract that favors the incumbent as it induces too little "trade" between the initial vertical structure and the entrant; there is socially too much foreclosure. Note that both our theory and theirs yield the same positive implication: the incumbent is favored at the reprocurement stage. In our model, poison pills, for instance, have *efficiency* justifications as well as Aghion-Bolton's *anti-competitive* motives.[26]

VIII. CONCLUDING REMARKS

In this article we have provided answers to some questions that

Williamson (1976) and the Chicago school placed on the research agenda concerning the optimal organization of franchise bidding for natural monopolies. We arrived at a relatively pessimistic assessment of the virtues of second-sourcing (or takeover) when substantial investments are at stake.[27] The incumbent should be favored at the reprocurement stage precisely when the investment is transferable, that is, when the second source is at not too much of a cost disadvantage. Indeed the incumbent should be favored more the higher the probability of second sourcing. To pursue this research it seems desirable to study various forms of non-commitment due either to incomplete contracting and renegotiation (see Grossman-Hart (1986), Klein et al (1978), Tirole (1986) and Williamson (1975) for investment concerns and Laffont-Tirole (1988b) for the ratcheting problem); or to the possibility of mutually advantageous renegotiation (see Dewatripont (1986), Hart-Tirole (1987) and Laffont-Tirole (1988a)).

In an incomplete contract setting, property rights do serve as switching incentives together with cancellation and entry fees. For instance, in defense procurement, the government sometimes owns the property rights to data and technological information and sometimes does not. Leaving the property right to the defense contractor can be viewed as a way of biasing the reprocurement stage in his favor for the government must bargain with and pay some money (the equivalent of a cancellation fee) to the defense contractor for the right to supply the relevant information to a second source. Property rights have thus some of the features of the switching incentives considered in this paper. In a takeover context, the corporate charger may influence the ease with which a raider can take control of the firm, through super-majority provisions and staggered board elections. In this respect, it is interesting to note that Grossman and Hart (1987) argue informally that family-run firms may sink considerable investments, and therefore may want to fight control changes through the allocation of voting rights.

Last, we showed that a rich yet tractable model can be built that yields testable equilibrium relationships between switching incentives (like golden parachutes and poison pills), managerial incentive schemes (like cost sharing and stock options), probability of second sourcing, and incumbent performance.

FOOTNOTES

1) See also Goldberg (1976), and in the context of the regulation of electric utilities, Joskow and Schmalensee (1983). An early reference on the topic is Pigou (1920) whose contribution is described in Goldberg (1981).

2) Williamson (1976) also mentions administrative and political incumbency advantages, which we will not study here.

3) Williamson also makes the important point that contracts are necessarily incomplete. Indeed, investment has been a major concern in the literature on the expropriation of relation-specific investment under incomplete contracting (Williamson (1975, 1985), Grossman-Hart (1986), Hart-Moore (1988), Green and Laffont (1988)). Although we also emphasize investment incentives, we depart from this literature in several important respects. First, we assume away unforeseen contingencies and analyze complete contracting. Second, the literature on incomplete contracting studies the role of ownership; we take ownership as given and analyze switching incentives. Third, whether the parties can contract on investment does not matter in the absence of second sourcing in our model while it does under incomplete contracting; assuming complete contracting allows us to focus on the effects of second sourcing in a cleaner way.

4) Assuming $\psi''' \geq 0$ makes stochastic schemes non-optimal (see Laffont and Tirole (1987)).

5) This assumption of monotone hazard rate prevents bunching in the static model. It is satisfied by most common distributions (for example, uniform, exponential, Pareto, logistic).

6) See Laffont and Tirole (1988a) (1988b) for analyses of procurement with renegotiation proof commitment and short-term commitments respectively.

7) Because of our assumption about two-period commitment we need consider only an intertemporal individual rationality constraint.

8) Because of the shadow cost of public funds, the firm's rent is set equal to 0 under symmetric information.

9) For $k=1$, the problem is always quasi-concave. For $k \neq 1$, a sufficient condition for quasi-concavity is that the marginal cost of the investment is high enough (see Laffont and Tirole (1987a)).

10) It should be noted that when one of the parties' preferences is not quasi-linear, an informed principal (regulator) strictly gains by not revealing his information to the agent (entrant) at or before the contract proposal stage. That is, by pooling at the contract proposal stage, the different types of principal (referring here to the possible values of β, by abuse of terminology) can trade the slack variables corresponding to the agent's individual rationality and incentive compatibility constraints. This may introduce a tension between first period efficiency and optimal regulation of the entrant. Also, the Maskin-Tirole result applies as long as the principal's information does not enter the agent's utility function; in particular, the agent's information can enter the principal's objective function, as is the case here.

An alternative proof of the fact that it is harmful to induce pooling in period 1 can be given as follows. From Laffont-Tirole (1987b) we know that the second period mechanism can be implemented by a dominant strategy (Vickrey) auction, which makes irrelevant the information transmitted by the observation of C_1. Then, pooling induces no gain in the second period, and it is harmful in the first.

11) We will use dots or full derivative notation for derivatives whenever a

prime might be misleading (primes are also used to index the entrant).

12) See Guesnerie and Laffont (1984).

13) If the marginal cost of investment is high enough, the first-order approach we take here in this moral hazard problem is valid.

14) Note also that common auctions for contract renewal belong to the class of mechanisms considered here.

15) For extensions of this result see Laffont and Tirole (1987b), McAfee and McMillan (1987), Caillaud et al. (1986), Melumad and Reichelstein (1986), Picard (1987), Rogerson (1987).

16) Apply the Lemma in the Appendix to equation (A.10).

17) These conditions are sufficient for $b^2\delta$ small.

18) $P(\beta)$ must satisfy: $U'(\beta^*(\beta))-P(\beta)=U_2(\beta)$.

19) There is, of course, a large diversity of ways to acquire firms, from friendly mergers to proxy fights. The view that managerial teams bid against each other may be a good first approximation and is taken in Blair et al. (1986), Grossman-Hart (1987) and Harris Raviv (1987), among others. It should be noted that other reasonable descriptions of the auctioning process would yield results similar to those we derive. For instance, suppose that the raider buys the whole firm, which then goes private. The auction is then equivalent to offering a fixed-price contract to the second source; the raider is made the residual claimant for the firm's second-period profit. Redoing the analysis by assuming that only a fixed-price contract can be offered to the entrant does not alter our intuitions.

Note also that our allowing discrimination among the raider's types yields the result that after a takeover, the firm goes private ($K'\simeq 1$) when the raider is very efficient (β close to $\underline{\beta}$) and does not when the raider is less efficient ($K'<1$ for higher βs).

20) A slight difference with our social planner formulation is that the shareholders do not care directly about the managers' welfare. But none of our qualitative results is affected by this change in the principal's objective function.

21) See Fudenberg-Tirole (1988) for a formal analysis of this issue in a model without takeovers.

22) Although the process of investing per se is likely to be observed by the market, the investment expenditure may not be straightforwardly derived from accounting data (recall Williamson's argument), and the quality of the investment may be hard to assess. In a similar spirit, Ruback (1986, p. 72) argues that "the management of most corporations has private information about the future prospects of the firm. This information usually includes plans, strategies, ideas, and patents that cannot be made public. Even if they are efficient, market prices cannot include the value of information that the market does not have." To the extent that plans, strategies, ideas, and patents result from investments, Ruback's argument fits with the notion that a non-negligible fraction of investment is not reflected in the market valuation of the firm. Hence, we believe that post-termination stock options, although they cannot be completely excluded, are at most a limited instrument to induce managers to invest.

23) Our view that the shareholders organize a bidding contest between managerial teams may be too simplistic so that caution should be exercised when applying our conclusions. But it is worth noting that Walking and Long [1986] found that managers with large stock holdings are less likely to oppose takeovers than managers with small stock holdings and that Malatesta and Walking [1986] provided evidence that firms who adopt poison pill defenses are relatively unprofitable. Such empirical evidence is consistent with our normative analysis.

24) It is worth recalling the intuition of why the golden parachute (respectively, the poison pill) should increase (respectively, decrease) with the manager's ability and performance. A first guess might have been that bad managers should be encouraged to leave with large golden parachutes and small poison pills. This, however, is not correct as bidding between managers already selects the best managers. Our point is that the auction should be rigged to encourage managers to invest. A (good) manager with probability .9 of keeping his job selects roughly the right amount of investment, and further incentives are not needed. A (bad) manager with probability .1 of keeping his job chooses an inefficiently low investment since with probability .9 this investment goes to a rival manager. A small golden parachute or a large poison pill increase his probability of keeping his job and his incentive to invest.

25) We suspect that the many shark repellents are far from being substitutes and involve fairly different social costs. It would be worthwhile to investigate how each favors the incumbent managerial team.

26) Hermalin (1987) and Stein (1988) also analyze the popular argument that the takeover threat may lead to underinvestment. Their models differ from ours in many respects and can be thought of as complementary. In these analyses, investment pays off before the raider enters the market for corporate control (i.e., in period 1 in the context of our model). Incumbent managers may not invest even in the presence of profitable opportunities because the probability of success of the investment is positively correlated with the manager's ability, and a failure signals a low ability and may encourage a takeover. These papers emphasize how signaling (managerial career concerns) distorts managerial *decisions* (more generally than investments) that might convey information about

managers. We assume, in contrast, that investment has long-delayed effects, and we focus on the intertemporal evolution of managerial profit-sharing schemes (closer in spirit to our work is the signal-jamming model of Stein (1987) in which a manager overborrows to mislead the market about his firm's worth).

27) We considered the case in which the number of potential entrants was exogenous (actually equal to one, but the analysis generalizes straightforwardly to any number). When this number is endogenous and increases with the expected rent to be gained in the second period auction, new effects should be taken into account. Favoring entrants now increases the competitiveness of the second period auction.

APPENDIX

Proof of Proposition 1:

The optimization problem is quasi-concave (for λ small enough) and separable. For given β and $\beta^*(\beta)$, we can maximize the inside integral with respect to C' under the constraints (16) and (17). The first-order condition of this control problem (see Laffont and Tirole (1987a) yields effort $e'(\beta') = e^*(\beta')$, where $e^*(\beta')$ is given by:

$$\psi'(e^*(\beta')) = 1 - \frac{\lambda}{1+\lambda} \frac{F(\beta')}{f(\beta')} \psi''(e^*(\beta')) \tag{A.1}$$

for any $\beta' < \beta^*(\beta)$. Equation (A.1) defines the optimal effort level and therefore the optimal C' function:

$$C'(\beta'/\beta) = \beta' - ki(\beta) - e^*(\beta'). \tag{A.2}$$

If he is selected $(\beta' < \beta^*(\beta))$, the entrant has rent:

$$V^*(\beta'/\beta) = \int_{\beta'}^{\beta^*(\beta)} \psi'(e^*(\tilde{\beta}))d\tilde{\beta}. \tag{A.3}$$

We can now maximize (19) with respect to $C_1(\beta)$, $C_2(\beta)$, $\beta^*(\beta)$ and $i(\beta)$ subject to (12), (13), and (15).

Let $\mu(\beta)$ (resp. $\nu(\beta)$) be the multiplier of the constraint (12) (resp. (15)).

The Pontryagin's principle yields:

$$\dot{\mu}(\beta) = -\frac{\partial H}{\partial U} = \lambda f(\beta) \tag{A.4}$$

Using the transversality condition at $\underline{\beta}$ we have:

$$\mu(\beta) = \lambda F(\beta). \tag{A.5}$$

Maximization with respect to C_1, C_2, β^*, i gives.

$$\psi'(e_1(\beta)) = 1 - \frac{\lambda F(\beta)}{(1+\lambda)f(\beta)} \psi''(e_1(\beta)) + \frac{\nu(\beta)d'(i)}{(1+\lambda)f(\beta)} \psi''(e_1(\beta)) \tag{A.6}$$

$$\psi'(e_2(\beta)) = 1 - \frac{\lambda}{1+\lambda}\,\frac{F(\beta)}{f(\beta)}\,\psi''(e_2(\beta)) - \frac{v(\beta)}{(1+\lambda)f(\beta)}\,\psi''(e_2(\beta)) \tag{A.7}$$

$$[\beta - i - e_2(\beta) + \psi(e_2(\beta))] - [\beta^* - ki - e'(\beta^*) + \psi(e'(\beta^*))] \tag{A.8}$$

$$= \frac{\lambda}{1+\lambda}\left[\frac{F(\beta^*)}{f(\beta^*)}\,\psi'(e'(\beta^*)) - \frac{F(\beta)}{f(\beta)}\,\psi'(e_2(\beta))\right]$$

$$- \frac{v(\beta)}{(1+\lambda)f(\beta)}\,\psi'(e_2(\beta))$$

$$0 = -(1+\lambda)f(\beta)d'(i)\psi'(e_1(\beta)) + (1+\lambda)\delta(1 - F(\beta^*(\beta)))\psi'(e_2(\beta))f(\beta) \tag{A.9}$$

$$- \lambda F(\beta)[d'(i)\psi''(e_1(\beta)) - \delta(1 - F(\beta^*(\beta)))\psi''(e_2(\beta))]$$

$$+ v(\beta)[d'(i)^2\psi''(e_1(\beta)) + \psi'(e_1(\beta)) + \delta(1 - F(\beta^*(\beta)))\psi''(e_2(\beta))]$$

$$+ f(\beta)k\,\delta[\int_{\underline{\beta}}^{\beta^*(\beta)} (1+\lambda)\psi'(e'(\beta'))f(\beta')d\beta' + \lambda\int_{\underline{\beta}}^{\beta^*(\beta)}\int_{\beta'}^{\beta^*(\beta)} \psi''(e'(\bar{\beta})d\bar{\beta}f(\beta')d\beta'].$$

Integrating the last line of (A.9) and using the first-order condition with respect to C' we can replace this last line by: $f(\beta)\delta kF(\beta^*(\beta))(1+\lambda)$. Using the other first order conditions, (A.9) reduces to:

$$v(\beta) = \frac{(1+\lambda)f(\beta)[d'(i) - \delta((1 - F(\beta^*(\beta)) + kF(\beta^*(\beta)))]}{\psi'(e_1(\beta))} \tag{A.10}$$

We first show that the following result holds:

Lemma: $v(\beta) \leq 0$ for any β.

Proof: Substituting (15) in (A.10) yields:

$$v(\beta) = \frac{(1+\lambda)f(\beta)}{[\psi'(e_1(\beta))]^2}\,\delta(1 - F(\beta^*(\beta)))[\psi'(e_2(\beta)) - \psi'(e_1(\beta)) \tag{A.11}$$

$$- k\,\frac{F(\beta^*(\beta))}{1 - F(\beta^*(\beta))}\cdot\psi'(e_1(\beta))].$$

Suppose on the contrary that $v(\beta) > 0$. From (A.11) and $\psi'' > 0$

$$e_2(\beta) > e_1(\beta). \tag{A.12}$$

From $\psi''' > 0$, we have

$$\psi''(e_2(\beta)) > \psi''(e_1(\beta)). \tag{A.13}$$

Let us now rewrite the first-order conditions relative to C_1 and C_2 as follows:

$$\psi'(e_1(\beta)) + \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e_1(\beta)) = 1 + \frac{v(\beta)d'(i)}{(1+\lambda)f(\beta)}\psi''(e_1(\beta)). \tag{A.14}$$

$$\psi'(e_2(\beta)) + \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e_2(\beta)) = 1 - \frac{v(\beta)}{(1+\lambda)f(\beta)}\psi''(e_2(\beta)). \tag{A.15}$$

If $v(\beta) > 0$, the left hand side of (A.14) is larger than the left hand side of (A.15) implying $e_1(\beta) > e_2(\beta)$, a contradiction.

We see from this lemma and (A.6) and (A.7) that $K_1 < K_2$ (since $\psi'(e_1(\beta)) < \psi'(e_2(\beta))$). The first period contract is closer to a cost-plus contract than the contract of the second period.

We show now that the incumbent is favored. Let

$$\Delta(\beta,\beta^*) \equiv (\beta^* - ki) - (\beta - i),$$

$$h(\beta,e_2) \equiv \psi(e_2) - e_2 + \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi'(e_2) + \frac{v(\beta)}{(1+\lambda)f(\beta)}\psi'(e_2)$$

and

$$g(\beta^*,e') \equiv \psi(e') - e' + \frac{\lambda}{1+\lambda}\frac{F(\beta^*)}{f(\beta^*)}\psi'(e').$$

Using (A.6) and (A.7), (A.8) can be written:

$$\Delta(\beta,\beta^*) = \max_{e_2}\{h(\beta,e_2)\} - \max_{e'}\{g(\beta^*,e')\}. \tag{A.16}$$

Equation (A.16) implies that $\Delta(\beta,\beta) < 0$, as $v(\beta) < 0$ implies that $h(\beta,e) < g(\beta,e)$ for all e. But the definition of Δ yields $\Delta(\beta,\beta) \geq 0$, a contradiction if $\beta^* = \beta$ were the solution. But

$$\frac{\partial}{\partial \beta^*}\,[\Delta(\beta,\beta^*) - \max_{e_2}\,\{h(\beta,e_2)\} + \max_{e'}\,\{g(\beta^*,e')\}] \tag{A.17}$$

$$= 1 + \frac{\lambda}{1+\lambda}\,\psi'(e')\frac{d}{d\beta^*}\left[\frac{F(\beta^*)}{f(\beta^*)}\right] > 0$$

Hence, for (A.16) to be satisfied, one needs $\beta^* < \beta$, which for $k = 1$ implies that the incumbent is favored.

When investment is observable or when $\beta = \underline{\beta}$, $\nu(\beta) = 0$. From (A.6) and (A.7), the incumbent's effort is identical in period 1 and 2. (A.8) then yields $\beta^*(\beta) = \beta$.

Q.E.D.

Proof of Proposition 2: From Laffont and Tirole (1986), we know that, under slightly stronger assumptions ensuring second-order conditions, we can rewrite the incentive contract of the entrant as a menu of incentive schemes which are linear in the cost overrun:

$$t'(C',\beta',\beta) = \tilde{G}\,(\beta',\beta) - \tilde{K}\,(\beta')(C' - C'^*(\beta'/\beta) \tag{A.18}$$

with $\tilde{K}\,(\beta') = \psi'(e'^*(\beta'))$ and $C'^*(\beta'/\beta) = \beta' - e'^*(\beta') - ki\,(\beta)$.

Denoting $C^{*\prime}(\beta') = C'^a$ and remembering the second-order condition $\dfrac{dC'^*}{d\beta'} \geq 0$.

$$t'(C',C'^a) = G\,(C'^a,C_1^a) - K\,(C'^a)(C' - C'^a)$$

with

$$K\,(C'^a) = \psi'(e^*(\beta')) = \psi'(e^*(\beta'^{-1}(C'^a)))$$

We can now extent the reasoning to the case of the incumbent. The transfer to the incumbent can be decomposed into two menus of linear incentive schemes, one for each period:

$$t_1(C_1,C_1^a) = G_2(C_1^a) - K_1(C_1^a)(C_1 - C_1^a)$$

$$t_2(C_2,C_2^a,C_1^a) = G_2(C_1^a,C_2^a) - K_2(C_2^a)(C_2 - C_2^a)$$

with

$$K_1(C_1^a) = \psi'(e_1^*(\beta)) = \psi'(e_1^*(\beta^{-1}(C_1^a)))$$

$$K_2(C_2^a) = \psi'(e_2^*(\beta)) = \psi'(e_2^*(\beta^{-1}(C_2^a)))$$

and the decomposition between G_1 and G_2 is arbitrary with a joint constraint (only their discounted sum matters, but one can choose G_1 and G_2 so that the I.R. constraint is binding in each period).

From (A.6), (A.7) and $\nu(\beta) < 0$ we see that $K_1 < K_2$ since $\psi'(e_1(\beta)) < \psi'(e_2(\beta))$. When investment is observable $\nu(\beta) = 0$ and $K_1 = K_2$.

Q.E.D.

Proof of Proposition 3

The intuition for the proof is that the incumbent fully internalizes the benefits of his investment when $k = 0$. Hence, the shadow price of the first-order condition for investment, $\nu(\beta)$, should be equal to 0 even when investment is unobservable. If $\nu(\beta) = 0$, then $e_1(\beta) = e_2(\beta) = e^*(\beta)$ from (A.6) and (A.7). (A.10) yields

$$d'(i(\beta)) = \delta(1 - F(\beta^*(\beta))). \tag{A.19}$$

and (A.9) is then satisfied. The functions $i(\beta)$ and $\beta^*(\beta)$ are defined by equations (A.8) and (A.19). Note that (A.8) can be rewritten as:

$$\beta^* - (\beta - i) = \max_e\{g(\beta, e)\} - \max_e\{g(\beta^*, e)\}\,. \tag{A.20}$$

where

$$g(\beta, e) \equiv \psi(e) - e + \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi'(e).$$

From the envelope theorem and the monotone hazard rate condition, the right-hand side of (A.20) is decreasing in β^* and is thus positive for $\beta^* < \beta$. Now suppose that $\beta^* \le \beta - i$. Then $\beta^* < \beta$ and thus the right-hand side of (A.20) is strictly positive, while the left-hand side is non-positive, a contradiction. We conclude that $\beta^* > \beta - i$.

Q.E.D.

Proof of Proposition 4

The first-order conditions of the regulator's quasi-concave maximization problem are:

$$\psi'(e'(\beta')) = 1 - \frac{\lambda}{1+\lambda}\frac{F(\beta')}{f(\beta')}\psi''(e'(\beta')) \tag{A.21}$$

$$\psi'(e_1(\beta)) = 1 + \delta a + \delta b(1 - F(\beta^*(\beta))) - \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e_1(\beta)) \tag{A.22}$$

$$\psi'(e_2(\beta)) = 1 - \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}(1-a-b)\psi''(e_2(\beta)) \tag{A.23}$$

$$\beta - be_1(\beta) - e_2(\beta) + \psi(e_2(\beta)) - (\beta^* - e'(\beta^*) + \psi(e'(\beta^*))) \tag{A.24}$$

$$= \frac{\lambda}{1+\lambda}\left[\frac{F(\beta^*)}{f(\beta^*)}\psi'(e'(\beta^*)) - (1-a-b)\frac{F(\beta)}{f(\beta)}\psi'(e_2(\beta))\right]$$

At $a = 0$, $\beta^* = \beta$. Differentiating (A.24) and using (A.21) and (A.23) gives $\frac{d\beta^*}{da} < 0$ for any a, which proves Proposition 4.

Q.E.D.

Proof of Proposition 6: We will assume that $\beta^*(\beta)$ and $v(\beta)$ are differentiable. This can be proved by using the implicit function theorem and the first-order conditions.

Let us first note that, $\frac{d\beta^*}{d\beta}(\underline{\beta}) < 1$. This is due to the fact that $\beta^*(\underline{\beta}) = \underline{\beta}$ and $\beta^*(\beta) < \beta$ for $\beta > \underline{\beta}$ (from Proposition 1).

Second, we know that $e' \equiv e'(\beta^*(\beta))$ is equal to $e_2 \equiv e_2(\beta)$ at $\beta = \underline{\beta}$ (from Proposition 1). Hence, at $\underline{\beta}$, one has $e_2 > e' \frac{d\beta^*}{d\beta}$. But differentiating (A.8) in the uniform-quadratic case, and using the first-order conditions (A.6) and (A.7) yields:

$$1 - \frac{d\beta^*}{d\beta} = \frac{\lambda}{1+\lambda}\left[e'\frac{d\beta^*}{d\beta} - e_2\right] - \frac{e_2}{1+\lambda}\frac{dv}{d\beta}. \tag{A.25}$$

This implies that at $\beta = \underline{\beta}$, $dv/d\beta$ is negative. Now

$$\frac{d}{d\beta}(e' - e_2) = \frac{de'}{d\beta'}\,\frac{d\beta^*}{d\beta} - \frac{de_2}{d\beta}.$$

Using (A.6) and (A.7) in the linear-quadratic case yields:

$$\frac{d}{d\beta}(e' - e_2) = \frac{\lambda}{1+\lambda}\left[1 - \frac{d\beta^*}{d\beta}\right] + \frac{1}{1+\lambda}\,\frac{dv}{d\beta} \tag{A.26}$$

But (A.25) implies that, at $\beta = \underline{\beta}$,

$$1 - \frac{d\beta^*}{d\beta} < -\frac{e_2}{1+\lambda}\,\frac{dv}{d\beta} \quad \text{so that}$$

$$\frac{d}{d\beta}(e' - e_2) < \frac{1}{1+\lambda}\,\frac{dv}{d\beta}\left[1 - \frac{\lambda e_2}{1+\lambda}\right] < 0,$$

as $e_2(\underline{\beta}) = 1$ and $1 > (\lambda/(1+\lambda))$. So, $e' < e_2$ in a neighborhood of $\underline{\beta}$. Now, consider the lowest $\beta > \underline{\beta}$ such that:

either A: $e'(\beta)\,\dfrac{d\beta^*}{d\beta}(\beta) = e_2(\beta)$

or B: $e'(\beta) = e_2(\beta)$

or C: $\dfrac{d\beta^*}{d\beta}(\beta) = 1.$

Condition A cannot be satisfied strictly before B or C is, as $\dfrac{d\beta^*}{d\beta} < 1$ and $e' < e_2$ in a neighborhood of $\underline{\beta}$.

Suppose that condition B is satisfied. So, at this β, $e'\dfrac{d\beta^*}{d\beta} \le e_2$ and $\dfrac{d\beta^*}{d\beta} \le 1$. The first inequality, together with (A.25) implies that

$$1 - \frac{d\beta^*}{d\beta} \le -\frac{e_2}{1+\lambda}\,\frac{dv}{d\beta}, \tag{A.27}$$

so that $\dfrac{dv}{d\beta} \le 0$. Using (A.26), and by the same reasoning as before, we obtain

$$\frac{d}{d\beta}(e' - e_2) \le \frac{1}{1+\lambda}\,\frac{dv}{d\beta}\left[1 - \frac{\lambda e_2}{1+\lambda}\right] \le 0 \qquad \text{(A.28)}$$

(as $e_2(\beta) = e'(\beta^*(\beta)) < 1$, from equation (A.1)). So the function $e' - e_2$ cannot become positive at β, as it is negative earlier and has a negative slope.

Last, suppose condition C is satisfied. Equation (A.25) can be rewritten as:

$$\frac{\lambda}{1+\lambda}(e' - e_2) = \frac{e_2}{1+\lambda}\,\frac{dv}{d\beta}. \qquad \text{(A.29)}$$

But (A.6) and (A.7) yield:

$$e' - e_2 = \frac{1}{1+\lambda}\,\frac{dv}{d\beta}. \qquad \text{(A.30)}$$

It is easy to see that (A.29) and (A.30) are inconsistent unless $e_2 = \frac{\lambda}{1+\lambda}$.[20] But, from (A.6), and the fact that $v(\beta) < 0$, $e_2 > 1 - \frac{\lambda}{1+\lambda}$. Since we assumed that $1 > 2\lambda/(1+\lambda)$, we obtain a contradiction.

Thus, neither of the three conditions, A, B, or C can obtain to the right of $\underline{\beta}$. This proves Proposition 6.

Q.E.D.

REFERENCES

Aghion, P. and B. Bolton (1987) "Contracts as a Barrier to Entry," *American Economic Review*, 77, 388-401.

Anton, J. and D. Yao (1987) "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement," *Rand Journal of Economics*, 18, 57-76.

Baron, D. and D. Besanko (1984) "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1, 267-302.

Blair, D., Gerard, J. and D. Golbe (1986) "Unbundling the Voting Rights and the Residual Profit Claims of Common Shares," Mimeo, Rutgers University.

Caillaud, B. (1985) "Regulation, Competition and Asymmetric Information," forthcoming, *Journal of Economic Theory*.

Caillaud, B., Guesnerie, R. and P. Rey (1986) "Noisy Observation in Adverse Selection Models," Mimeo, INSEE.

Demsetz, H. (1968) "Why Regulate Utilities?", *Journal of Law and Economics*, 11, 55-66.

Demski, J., Sappington, D. and P. Spiller (1987) "Managing Supplier Switching," *Rand Journal of Economics*, 18, 77-97.

Dewatripont, M. (1985) "Renegotiation and Information Revelation over Time in Optimal Labor Contracts," also Ch.1 in *On the Theory of Commitment, with Applications to the Labor Market*, Ph.D. Thesis, Harvard University.

Fudenberg, D. and J. Tirole (1988) "Moral Hazard and Renegotiation in Agency Contracts," MIT, DP494.

Goldberg, V. (1976) "Regulation and Administered Contracts," *Bell Journal of Economics*, 7, 426-452.

Goldberg, V. (1981) "Pigou on Complex Contracts and Welfare Economics," *Research in Law and Economics*, 3, 39-51.

Green, J. and J. J. Laffont (1988) "Investment and Renegotiation," Mimeo, Harvard University.

Grossman, S. and O. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691-719.

Grossman, S. and O. Hart (1987) "One Share/One Vote and the Market for Corporate Control," Working Paper 440, MIT.

Guesnerie, R. and J. J. Laffont (1984) "A Complete Solution to a Class of Principal–Agent Problems with an Application to the Control of a Self–Managed Firm," *Journal of Public Economics*, 25, 329-369.

Harris, M. and A. Raviv (1987) "Corporate Governance: Voting Rights and Majority Rules," Working Paper 33, Kellogg GSM, Northwestern University.

Hart, O. and J. Moore (1988) "Incomplete Contracts and Renegotiation," *Econometrica*, 56: 755-786.

Hart, O. and J. Tirole (1987) "Contract Renegotiation and Coasian Dynamics," forthcoming in *Review of Economic Studies*.

Hermalin, B. (1987) "Adverse Effects of the Threat of Takeovers," Mimeo, MIT.

Holmström, B. and J. Tirole (1987) "The Theory of the Firm," forthcoming in R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, North Holland.

Joskow, P. and R. Schmalensee (1983) *Markets for Power*, MIT Press.

Klein, B., R. Crawford and A. Alchian (1978) "Vertical Integration, Appropriable Rents and the Competitive Contracting Process," *Journal of Law and Economics*, 21, 297-326.

Laffont, J. J. and J. Tirole (1986) "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614-641.

Laffont, J. J. and J. Tirole (1987a) "Repeated Auctions of Incentive Contracts, Investment and Bidding Party, with an Application to Takeovers," MIT DP463.

Laffont, J. J. and J. Tirole (1987b) "Auctioning Incentive Contracts," *Journal of Political Economy*, 95, 921-937.

Laffont, J. J. and J. Tirole (1988a) "Adverse Selection and Renegotiation in Procurement," CALTECH discussion paper.

Laffont, J. J. and J. Tirole (1988b) "The Dynamics of Incentive Contracts," forthcoming *Econometrica*.

Malatesta, P., and R. Walking (1986) "Poison Pill Securities: Stockholder Wealth, Profitability, and Ownership Structure," Mimeo, University of Washington and Ohio State University.

Maskin, E. and J. Tirole (1986) "The Principal–Agent Relationship with an Informed Principal: I: Private Values," Mimeo, MIT.

McAfee, P. and J. McMillan (1984) "Discrimination in Auctions," Mimeo, University of Western Ontario.

McAfee, P. and J. McMillan (1987) "Competition for Agency Contracts," *Rand Journal of Economics*, 18, 296-307.

Melumad, N. and A. Reichelstein (1986) "Value of Communication in Agencies," Mimeo, GSB, Stanford University.

Myerson, R. (1981) "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58-73.

Picard, P. (1987) "On the Design of Incentive Schemes under Moral Hazard and Adverse Selection," *Journal of Public Economics*, 33, 305-332.

Pigou, A. C. (1920) *The Economics of Welfare*, McMillan.

Posner, R. (1972) "The Appropriate Scope of Regulation in the Cable Television Industry," *Bell Journal of Economics*, 3, 98-129.

Riordan, M. and D. Sappington (1987) "Awarding Monopoly Franchises," *American Economic Review*, 77, 375-387.

Rogerson, W. (1987) "On the Optimality of Menus of Linear Contracts," Mimeo, Northwestern University.

Ruback, R. (1986) "An Overview of Takeover Defenses," in A. Auerbach editor, *Mergers and Acquisitions*, NBER prepublication report.

Scharfstein, D. (1988) "The Disciplinary Role of Takeovers," *Review of Economic Studies*, LV: 185–200.

Stein, J. (1988a) "Takeover Threats and Managerial Myopia," *Journal of Political Economy*, 96, 61-80.

Stein, J. (1987) "Efficient Stock Prices, Inefficient Firms: A Signal–Jamming Model of Myopic Corporate Behavior," Mimeo, Harvard Business School.

Stigler, G. (1968) *The Organization of Industry*, Homewood, III: Richard D. Irwing.

Tirole, J. (1986) "Procurement and Renegotiation," *Journal of Political Economy*, 94, 235-259.

Walking, R. and M. Long (1984) "Agency Theory, Managerial Welfare, and Takeover Bid Resistance," *Rand Journal of Economics*, 15, 54-68.

Williamson, O. (1975) *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press.

Williamson, O. (1976) "Franchise Bidding for Natural Monopolies—in General and with Respect to CATV," *Bell Journal of Economics*, 7, 73-104.

Williamson, O. (1985) *The Economic Institutions of Capitalism*, New York: The Free Press.

THE ECONOMICS AND PRICING OF BROADBAND

Stanford L. Levin
Chairman and Professor of Economics
Southern Illinois University
at Edwardsville
Edwardsville, IL 62026-1102
618-692-2592

This paper was prepared for presentation at the Sixteenth Annual Telecommunications Policy Research Conference, Airlie House, Airlie, Virginia, October 30 - November 1, 1988.

THE ECONOMICS AND PRICING OF BROADBAND

I. Introduction

Broadband services delivered over fiber optic telecommunications networks are likely to change the entire telecommunications industry fundamentally. Not only may existing services, including telecommunications and television, be delivered differently, but new services not yet envisioned may become common. To date, much of what has been written about broadband has concerned technology only. Planning for broadband implementation requires considering broadband in the marketplace, including pricing scenarios.

The policy response to broadband is undoubtedly important, as the broadband that results will affect the structure, service availability, and pricing of telecommunications and related services in the coming years, as well as the international competitiveness of U. S. telecommunications services. Recent improvements in telecommunications technologies have emerged faster than previous ones, as seen clearly in switching, and this suggests that narrowband ISDN as envisioned today may already be obsolete. It is important, therefore, to plan ahead actively, either by deciding to let the market control developments following technology and demand, or, if necessary, by pursuing a more active regulatory policy.

This paper begins with a brief introduction to broadband technology. Enough of the technology is presented to provide an understanding of the resulting economics. This leads to important conclusions about capacity, marginal cost, and natural monopoly conditions which may result from broadband technology.

A review of microeconomic pricing theory suggests how total costs would be covered in a competitive industry with the cost structure of broadband, and some specific industry examples are available. This gives an understanding of broadband pricing not only under competition, but also under regulation, if it continues, which will recover total costs and which will parallel the pricing expected in a competitive market.

The paper concludes with a discussion of several important policy issues, including whether or not broadband services should be regulated, who will pay for these service, what tariffs might look like, and who will receive broadband services. This analysis, coupled with lessons from recent experience in telecommunications, leads to some recommendations for the industry and for policymakers to insure that the economically justified benefits of broadband technology are realized.

II. Technology

There are numerous excellent descriptions of broadband technology available [for example, Harrold and Strock, 1987]. For the purposes of pricing, however, it is important to remember that broadband is a delivery technology and not a service. The demand is for these services, not for broadband itself, and, indeed, broadband can be used to deliver current narrowband services as well as new services which may only be practical with broadband.

Basically, broadband, or bandwidth, is the speed of transmission, and high transmission speeds require fiber. With the economies of scale of installing fiber, relative to demand, and the ever-increasing ability to increase transmission speed, there appears to be the prospect for the elimination of any meaningful capacity constraint for transmission, especially for local exchange telecommunications companies. In addition, broadband technology continues to improve. For example, Bellcore has developed digital coding to compress a high definition television (HDTV) signal onto a single broadband channel, providing a 10:1 bandwidth reduction [Wilson, 1988].

At the same time, one must be careful extrapolating very far into the future. After all, not many years ago it would have been difficult to envision the uses being made of today's telecommunications network. It is also possible that while transmission capacity may be freely available, other parts of the network such as switching may have real capacity constraints associated with them. In any case, for the near future with broadband transmission, economies of scale appear important and may once again emerge to control the structure of local exchange telecommunications, which could exhibit some of the characteristics of a legitimate natural monopoly.

In a simple sense, the high transmission speed is accomplished by sending frames over the fiber more and more rapidly. Each frame consists of a header, giving customer and routing data for the frame, and the information part of the frame. For pricing, it is important that customers cannot readily share a frame but that the broadband provider knows how full the information portion of the frame is.

While technologists are working hard to develop various broadband technologies, strong commercial incentives are in place for the technology to become standardized. Broadband standardization is important and is developing, for example, with SONET [Warr, 1988; Boehm, 1988]. While a lack of standardization is certainly cause for concern, and while solutions may be difficult, standardization is likely to be worked out before it is an ultimate barrier to the introduction of broadband technology.

There are too many Federal Express type examples to rule out the likelihood that the availability of broadband will create new demands, at the same time that demands may spur the development of new services as customers imagine new possibilities. It may not be realistic, therefore, to expect accurate or even reasonable forecasts of demand; the real economic issue is who will bear the risk of developing and implementing broadband and if private investors are willing to accept this risk. Once the technology looks assured, the market will determine whether to go forward with broadband.

III. Pricing in Competitive Markets

An understanding of how prices are set in competitive markets helps to envision how broadband pricing may evolve, either in a regulated or non-regulated environment. This is in itself useful, as speculation on pricing helps companies, investors, and regulators understand the future of broadband, the opportunities for successful investment, and the need, if any, for active policy intervention or regulation.

Economics textbooks are full of explanations of how a firm making one product at one location and selling in one market will determine its price. While this analysis is instructive, it is not directly relevant to most firms in the U. S. economy which often produce more than one product or service, sometimes at more than one location, and sell in more than one market. This is certainly the situation for telecommunications companies and for potential providers of broadband facilities or services.

A firm producing more than one product or service usually incurs joint and common costs. Certainly the firm will not want to set the price of any one of its products below its marginal cost. At the same time, the firm must cover its total costs if it is to stay in business in the long run, and setting all prices at marginal cost may not provide enough revenue to do this. Consequently, some or all of the firm's prices must be above marginal cost. The extent to which different prices are greater than marginal cost depends on conditions in the markets for each of the different products or services. These conditions involve primarily elasticity of demand; the value of the product or service or its importance, for example, are not major determinants of the margin by which price exceeds marginal cost. The greater the elasticity of demand, the less price can exceed marginal cost in the market. The smaller the elasticity of demand, the more price can exceed marginal cost in the market. While this is nothing but the inverse elasticity rule or Ramsey pricing, it is nonetheless worth repeating.

When there is competition in a market, the market itself will ultimately determine the price; regulation, if it is present, may have some effect in the short run but will play little if any role in the longer run, as competitive forces dominate. If there is not competition in the market for the service, regulators may play a greater role in determining price. Economic efficiency will be greatest, and the public will benefit the most, however, to the extent that Ramsey prices are implemented.

Of course, this is nothing but price discrimination, which, despite the sometimes bad connotation, is widely practiced in competitive markets. Although regulators often profess to be guarding against price discrimination, most regulatory statues prohibit only "undue" price discrimination, and, in fact, regulated prices usually incorporate some form of price discrimination.

While airlines may be among the most sophisticated practitioners of price discrimination, perhaps out of necessity as marginal costs are low relative to average costs, examples are easy to find in many other markets. Like airlines, however, the marginal cost of using broadband to deliver services may be relatively low, so it is important to search for pricing schemes which will recover total costs. The challenge is to find a mechanism for engaging in successful price discrimination, just as is done in competitive markets. Airlines use advance purchase and length of stay, for example; if broadband pricing is to be successful, something comparable will be required.

IV. Pricing Broadband

As the technology becomes clearer and standardizes, the important broadband issue will be pricing. While there is certainly some interaction between availability and demand, broadband will still have to pass some market test. Except for the underlying marginal costs, however, pricing can be separated from technology. A failure to do so, or reliance on average costs for guidance, will result in unrealistic, inefficient, and perhaps unsustainable pricing scenarios which may inhibit the diffusion of broadband.

It is first worthwhile looking at the type of pricing that will almost certainly not result. Today's pricing rules, or something similar attempting equal charges for equal use of capacity, will not be applied in a way that yields either free local telephone service so that video-on-demand can be priced competitively or movies costing several hundred dollars so that local exchanging pricing can be maintained at current levels. Similarly, flat rate pricing which significantly raises the minimum access cost above today's local phone tariffs will not result in the market, nor will it be allowed by regulators (and, in any case, it is probably not very efficient).

Planning for broadband pricing should also not focus unduly on HDTV. Broadband will most likely begin by delivering existing narrowband services and may have to prove itself with these. HDTV and perhaps video-on-demand may well follow. At the same time, there will be new services, such as the ability to use video-on-demand to preview vacation destinations, hotels, and restaurants. The history of the electric utility and telecommunications industries shows that what they are providing now bears little resemblance to what they did initially; certainly the first local telephone service is far from widespread long distance and international voice and data services, dial-up information services, and facsimile transmissions. The broadband pricing strategies considered must incorporate this type of dynamic change if they are to provide any useful insight.

Many or most of the costs associated with broadband are fixed, so that marginal costs are zero or close to zero for transport. Marginal costs for switching and other aspects of providing broadband are less clear, especially outside of the shortest run, and marginal costs may not be zero, although they are still probably very low. While regulators talk extensively about setting prices equal to cost, this may be more talk than accomplishment in some cases, with certain prices regularly below marginal cost. In any case, it is time to move away from a strict "price equals cost" formula which may sound like a good rule but which has little foundation in economics for the way it is actually applied.

The appropriate model for recovering total costs for a multi-product firm is found from the examination of competitive firms. The basic rule is that all prices should be at least equal to marginal cost. (This will also insure that there is no price discrimination, the bane of regulators.) This is not, however, an excuse to use Ramsey pricing blindly, and incorrectly, to place much of the fixed costs on residential consumers, for example.

The pricing issue is more complex than it might seem at first. For example, at some point it is less expensive to provide tone dialing service to all customers than to offer both tone and pulse service. While initially there may have been a marginal cost to charge to tone dialing customers, at some point these customers are no longer marginal, and the extra cost is actually for providing pulse dialing. At this point, tone dialing should be standard, and the marginal cost is zero. Similarly, call waiting might sufficiently reduce switching congestion by reducing redialing so that total costs would fall rather than increase; the marginal cost is zero or negative, and call waiting is included as part of access.

Given the cost structure of broadband, the question, then, is how to use price discrimination successfully to recover the total costs of broadband while keeping all prices at least equal to marginal cost. The major implication of broadband technology is that such price discrimination can no longer be accomplished by type of use, as it sometimes is today. The telecommunications company no longer knows, for example, whether the customer is transmitting voice or data.

Price discrimination under broadband can, however, be accomplished by quantity. Pricing can vary with the number of frames, how full the information portion of the frame is, the complexity of the overhead and how full the overhead portion of the frame is, if the provider puts the information and routing on the frame overhead, and if the switch is used. Even though there may be cost considerations in handling different frames and in using the switch, there should still be ample opportunity for recovering joint and common costs. In addition, declining block pricing and other volume discounts may be used, and it may be possible to have different access charges for different types of customers.

For example, regulation in the electric utility industry has provided reduced prices for cogenerators and large customers with choices. Water companies with large reservoirs and no effective capacity constraint discriminate by type of customer to maintain low prices for small customers, rather than charging everyone the same flat rate, so that everyone can afford service. The only restriction is that the cost of alternatives is greater than the prices charged. If prices are determined in competitive markets, the markets will approximate Ramsey pricing. If the prices are determined by regulators in markets that are sufficiently non-competitive so that the regulators have some effect, approximating Ramsey pricing will, of course, maximize economic efficiency.

In addition, broadband providers may be able to charge for what is now free. For example, customers may pay for an HDTV signal, just as they now pay for cable to receive over-the-air signals, presumably of better quality or more easily. Based on this, customers will demand HDTV quality as it becomes available on videos for VCRs, and if they are willing to pay for current cable TV, they will certainly be willing to pay for HDTV.

Another form of price discrimination can be seen clearly from the early market for TVs and VCRs. Those customers with a greater willingness to pay make their purchases earlier and pay higher prices, contributing more to the recovery of fixed costs or up-front investment. Prices fall over time and the market expands. Similar pricing may help to recover broadband investment.

There are some important lessons from past telecommunications and other regulation. Handicapping, entry blocking, pricing inefficiencies, and regulation of technology have not really worked to the customer's advantage. Also, pricing should not focus too extensively on the issue of stranded investment and inadequate depreciation. Telecommunications companies should not go the route of the steel industry. AT&T is still considering additional write-offs, and the BOCs, who are increasing depreciation charges, may also take write-offs in the future. These are lessons to be reviewed when broadband is implemented.

Pricing broadband services along these lines will provide the best chance of recovering the investment in broadband, while the price structure will eventually extend broadband to more customers. Efficient pricing of broadband network services will also provide the proper basis for the intelligent network to compete with intelligent terminals, and it will not be necessary to select one or the other in advance.

A certain amount of price discrimination, coupled with pricing dependent on the characteristics of broadband use and volume discounts, is the type of pricing competitive markets suggest. Such pricing should give the best chance for success in the market for broadband, and it helps companies, regulators, and investors to understand the risks and problems they may confront as broadband technology continues to develop and is implemented.

V. Policy Implications

Installing fiber to the home is not an appropriate policy objective; rather, the objective is the provision of market-justified services. Even in this context, though, policy may be needed for the provision of infrastructure. At the same time, policy or regulation may place too many roadblocks which will lead to inefficient outcomes. The many issues raised by broadband deployment concern many aspects of telecommunications and related industries.

The initial incentive to develop broadband was probably in response to perceived needs of businesses or other large users. The broad public and residential customers will probably get broadband services only through the public network. It would be a mistake, however, to fail in the attempt to implement broadband by insisting that it must be available initially for everyone or for no one. Broadband must be deployed as it is justified in the market. A ubiquitous system will eventually develop, although this may take some time, as some aspects of the current telecommunications network are still not available everywhere.

The impending deployment of broadband is yet another pressing impetus for fixing local exchange pricing. For example, if local exchange prices are below cost, as is generally the case, and broadband services at at or above marginal cost, broadband may not be able to compete with current local exchange pricing, either in the market or in the regulated arena, and broadband will fail. Broadband services may not be able to command a sufficient premium above below-cost local exchange services, and regulators, if they are involved, will block broadband so that local exchange prices will not increase. It seems that improved local exchange pricing may be a prerequisite to widespread broadband deployment.

One of the most contentious issues surrounding broadband is who will provide the broadband delivery system. The answer depends on both economics and regulation. There may or may not be more than one broadband delivery system, depending on ultimate costs, economies of scale, and local market characteristics. Similarly, there may or may not be private networks supplementing all or part of the public network. Larger markets are more likely to have more than one system, particularly if absolute costs are low, while smaller markets, without one or a few large customers, will almost certainly have only one system. Both telecommunications companies and cable TV companies, along with others, are positioning themselves to provide the broadband network. Local exchange companies are interested because of their current services, while cable companies are interested because they currently supply television/video services.

Relaxing the cable/telco cross ownership ban may be central to an efficient solution to the broadband ownership conflict. The FCC is considering recommending a relaxation of this ban to Congress, voting to consider the issue July 20, 1988, and issuing a proposal for comments September 22, 1988 [Mason, 1988]. This proposal mentioned particularly a waiver for construction of integrated broadband systems, allowing telcos to provide cable and giving them an incentive to provide fiber to the home.

Telcos and cable companies each bring something to a prospective merger. Cable companies have customers and the right to a market for TV-type services. Telcos have large size (so do some cable companies), access to capital, better customer relations, and technological expertise.

While allowing potential competitors to merge is not usually good public policy, in this case it will remove an important impediment to broadband deployment. The structural conflicts make achieving a broadband network difficult, and the alternative is often not two networks but rather none. The question is how to get the best single network. At the same time, it may be possible to replace a cable monopoly with a better alternative.

Any evaluation of competition must also make the distinction between a competitive structure and competitive behavior, and competitive behavior in this case can be preserved. The owner may be required to operate the network as a common carrier. This may happen automatically, with pressure to sell the substantial excess network capacity. The owner's economic interest in exclusion should be low, and antitrust may provide a sufficiently effective enforcement threat and remedy.

Currently telcos are prevented by the MFJ from originating information or video services. They can, however, probably control a small number of the total channels without any adverse effect on competitive behavior; indeed, they are an important source of potential competition. Developing ONA experience may suggest how to resell the network and network services and how to get competitive behavior with one underlying network.

Even so, some regulation may be needed. This regulation will have to address whether or not BOCs and independent telcos involved in broadband and cable/telco cross ownership should be treated the same. They may or may not be allowed to originate programming. Regulation may also have to determine whether the broadband network provider should formally be required to adhere to common carrier status or whether there are sufficient market forces to make such regulation unnecessary. Recently, telcos have not upheld the concept of common carriage with their quick cave-ins on dial-a-porn and related issues, and this weakens the arguments for no regulation.

Relaxing the ban on cable/telco cross ownership can prevent unjustified competing systems, while providing maximum choice, better service, and lower prices to customers. Such accommodations are particularly likely to result as existing cable systems reach the end of their useful life; the opportunity will be excellent for a cable/telco combination to install a new multipurpose fiber-based broadband delivery system.

The regulatory infrastructure must be compatible with the new technology, but rate of return regulation appears incompatible with the development of ubiquitous broadband service. For example, traditional regulation has worked by control of entry, prices and the deployment of new technology. It is no longer possible to do this, as technological change is too rapid and market conditions change too frequently. In addition, the current regulatory process is titled against risky investments, such as broadband, the inadequate current depreciation rates can no longer be used to move cost recovery over time. Frequent state/federal conflicts may slow or prevent the development of broadband; current disputes may prevent getting necessary regulatory decisions and will certainly increase the time, cost, and uncertainty of getting these decisions.

The emergence of broadband forces a serious consideration of whether or not to regulate and if so, how to do so. Traditional regulation, and even price regulation, should not be necessary if behavior is competitive. In addition, before opting for regulation, policymakers should consider the extent to which regulators of broadband would be asked to mediate among competitors rather than between stockholders and customers. With broadband, regulation may be unnecessary and may, if used, find itself in an uncomfortable environment for which it is poorly prepared.

Price regulation may be the preferred alternative for broadband if regulation is necessary. It will maximize competition by removing entry barriers. Price regulation also will eliminate the guaranteed success for providers of new technology. Moving away from the more-or-less risk-free capital recovery of traditional regulation will force providers to base their decision on future market conditions and to write off sunk costs when appropriate. Whatever regulatory scheme is selected, including a reliance on markets, it should be the approach with the least regulation to accomplish public policy goals.

Broadband is likely to transform the entire telecommunications industry. The discussion must now begin to move from the technology to the availability of services, market demand considerations, and policy to insure that society receives the advantage of broadband services and that the U. S. remains competitive in world telecommunications. Competitive markets suggest some practical ways in which broadband can be priced, and an understanding of likely pricing, and of winners and losers, will help the debate on formulating public policy for broadband.

REFERENCES

Boehm, Rodney J., "SONET: An International Standard," Telecommunications, March, 1988, pp. 73-76.

Harrold, Daniel J., and Strock, Robert D., "The Broadband Universal Telecommunications Network," IEEE Communications Magazine, Vol. 25, No. 1., January, 1987, pp. 69-79.

Mason, Charles, "FCC Releases Cable/Telco Proposition," Telephony, September 26, 1988, Late-Breaking News.

Warr, Michael, "SONET -- Preparing for Phase II," Telephony, Transmission Special, October 1988, pp. 28-9.

Wilson, Carol, "Bellcore Devises Digital Coding for HDTV," Telephony, September 5, 1988, pp. 8-9.

RATINGS POLICY AND PUBLIC POLICY

Peter V. Miller
Department of Communication Studies
Northwestern University

DRAFT: NOT FOR CITATION WITHOUT PERMISSION

Prepared for the Airlie House Telecommunications Policy Conference, October 31, 1988.

> "Our concern was with the accuracy requirements appropriate to the kinds of decisions that are made on the basis of the ratings...[but] for the most part, we must acknowledge little clarity of understanding in this area...In general, users of statistical data do not formulate explicit accuracy requirements."
>
> Madow Committee, 1961.

A paradox of measurement theory and practice is that measures are evaluated according to how they are used, but users often have little say or interest in how measures are constructed. When the Madow Committee of the American Statistical Association evaluated broadcast ratings methods at the behest of Congress, they surveyed ratings data users to try to establish evaluation criteria. Those advertisers and broadcasters who replied gave them few clues. The committee also sought evidence of methodological research on ratings by the primary data users. They found little or none. The user community, theoretically the arbiter of ratings data quality, largely did not play that role.

And totally absent from discussions and research on ratings data quality were government policy makers. The Madow Committee did not include the FCC or any other government agency in its sampling frame for the user study. While Congress had mandated the evaluation of ratings methods, government agencies that could employ ratings data were not surveyed concerning their possible uses of such information, or the accuracy requirements for information upon which their decisions might be based. Had the FCC been included, the responses from the agency probably would not have been any more enlightening than those from broadcasters and advertisers. At that juncture, the agency might not have envisioned the day when ratings data would become the basis of policy. Further, as the committee noted, few users of statistical data attempt to establish accuracy requirements for the information they rely upon.

In the years since the Madow Committee report, and partly due to its recommendations, primary users of ratings data have become much more involved in issues of data quality. Thanks to the efforts of the National Association of Broadcasters, the Committee on Network Television Audience Measurement, the Committee on Local Television and Radio Audience Measurement, The Electronic Media Rating Council, the Advertising Research Foundation and others, we know much more about ratings data and its limitations now.

But governmental policy makers are still largely "out of the loop" in discussions about and developments in ratings methodology. As Webster (1988) notes, rating services do not take on governmental clients because the data would enter the public domain once it reached the agencies. Government policy interests benefit from industry clients' oversight efforts to the extent that the industry's priorities coincide with the information needs of policy makers. Policy interests suffer when the industry adopts a measurement course with different aims.

I argue in this paper that the industry's course has always diverged somewhat from that of the policy maker, and that the gap between what the industry measures and what the government requires is likely to widen. Policy makers use ratings data more because they are there than because they are suitable. The suitability of policy applications of ratings data is further challenged by recent major changes in the ratings business and some likely future developments. Most of my observations will concern television audience measurement, with only a few remarks about radio.

The Syndicated Research Context

Broadcast rating studies are one of a set of research enterprises that are often termed "syndicated studies." Syndicated research is characterized by its appeal to many clients with diverse (or "vague," viz. Spencer and Moses, 1985) interests, but who receive one common standardized report. For the research company, syndicated studies are much to be preferred over "custom" work, since they provide a more constant income stream and since a sizable profit can be made from the surplus clients recruited above the number needed to break even. Further, syndicated studies often provide the potential for special analyses of the core data base, for which extra fees can be charged. Finally, custom research work can "spin off" the syndicated study if recontacts with the main study sample are made and the resulting new data are merged with the original data set. For example, both Simmons Market Research Bureau and Mediamark Research Incorporated do large syndicated studies of media exposure and product purchasing. They then offer to recontact subsets of their large samples to collect more specialized information that can be merged with the data from the original contact.

Syndicated research, then, can be a highly profitable venture for the research firm because, once the investment in the initial data collection is recouped, the resale of the data in various forms provides substantial income. Marketing the data in different packages in an attempt to suit the interests of individual users is a task that occupies many research firm employees. Their sales efforts contribute to the impression that syndicated data are sufficiently flexible to address a wide array of questions. Moreover, marketing personnel wield considerable power in the corporate hierarchy, setting the specifications for the original data collection and negotiating with research practitioners within the firm when conflicts arise over what data packages can be sold and what packages are technically advisable.

Conflicts of this sort are common. The value of syndicated data to the research firm is maximized when it can be viewed by clients as both a standardized measurement product and a customized one. Marketing's job is to point out all the possibilities. But, from the research point of view, no study is designed to meet every purpose, and the quality of the the research is judged on how well it suits particular, well-defined purposes. Thus, research practitioners sometimes feel compelled to oppose the sales efforts of their marketing colleagues. Sometimes marketing's ideas are scotched, and others go forward with prominent disclaimers about the appropriateness of the data for the client's question.

The tension between what a study purports to measure and what it can be sold as measuring is particularly salient for syndicated research. Because of the standardized, large scale approach of syndicated studies, they come to be viewed as "official," and possibly more objective than custom research, as Webster (1988) has noted. The disinterested image of the research is at once a major appeal and a major drawback. Syndicated research appeals because it appears not to have particular interests in mind; for the same reason, it often lacks the ability to address directly an individual client's information needs.

Clients try to shape the design of the syndicated study to fit their own interests. This is particularly true of clients whose product is measured by the service -- e.g. broadcasters and publishers. Their efforts present a delicate problem for the syndicated measurement service, since acquiescing to the demands of one client may disadvantage others. Also, the cost structure of the syndicated study dictates that some design options are out of the question without a change in the rates charged to clients. This decision, again, is a touchy one. A former research director of MTV once told me that Nielsen should make design alterations to include viewing alternatives such as the cable network within the standardized package of national ratings and absorb the costs "for the good of the industry." This argument was not persuasive to Nielsen, and may partly explain MTV's early support of Nielsen's competitor, AGB. While Nielsen did not meet the demands of this particular client, syndicated services are generally more solicitous of clients when there is competition from other measurement services. The people meter "revolution" would not have happened if AGB had not entered the US market to challenge Nielsen.

In any event, the shape of a syndicated study is the result of decisions by the measurement service on which client information needs belong in the standardized package, and which must be covered by ad hoc research efforts with their own additional price tag. Clients will try to make their own interests the "industry's interests" to make sure that their products are measured, and/or to try to configure the measurement to their own advantage.

What is measured by ratings data?

The product of measurement service decisions about what it can and should measure and what it can and should sell, coupled with various client and competitive pressures, produces a set of numbers utilized and cited by clients and non-clients (such as government policy makers) for multifarious purposes. In the case of television ratings, the numbers purport to measure program audience delivery. Operationally, at the national level the various competing forces which have shaped rating numbers have agreed upon continuous, passive metering of set tuning as a primary component of the measure, and continuous human recording of individuals in the audience as a secondary component.

Meters. An early codification of the "tuning" emphasis is found in the Advertising Research Foundation's "recommended standards" for TV program audience size measurements (Advertising Research Foundation, 1954). This report offered an operational definition for "viewing" (set tuning), and suggested a unit of measurement (the household). Further, the ARF committee suggested that the rating services report the "average instantaneous audience" for a program (the average number of households tuning during each quarter hour), in preference to the "total audience" (the cumulative number of households tuning to some minimal part of a program).

Continuous, passive measurement at the household level provided information for an important use of ratings data: the buying and selling of small units of broadcast time. Multi-set households were rare, and viewing was conceived to be a family activity, done at home. Further, demographic information had not yet been elevated to an essential status by advertisers. Finally, electronic measurement had an appearance of impartial reliability that human self-report efforts lacked. Thus, household tuning became the dominant operationalization of television viewing.

Of course, as in any syndicated survey situation, not all interests were served ideally by this solution. Continuous household tuning measurement of viewing won out over alternative operational definitions that seem reasonable, and by some criteria, better. Tuning ignores the psychological aspects of viewing -- attention, perception, memory, affect, involvement, etc. -- which are conjured up when one thinks of the term, "watching." These mental processes were important parts of measurement approaches of several Nielsen competitors, but were irrelevant to the workings of Nielsen's audimeters.

In addition to sounding the death knell for several Nielsen competitors, the household tuning definition also ignored some potential uses of audience data. For example, depending on one's purposes, there might be a need for more than mere head (or household) counts; affect, involvement, and so forth might be important. Or, one might want more evidence of "viewing" than mere tuning -- some proof of attention might be useful. Such conceptualizations of audience delivery were relegated to the "unofficial" realm of custom research. Because they were not a part of the syndicated package, evidence concerning audience likes, dislikes, involvement, etc. would not be widely or regularly available to analysts of television audiences. Instead, measurement of these phenomena would be sporadic, unstandardized and often proprietary. Though they are not clients, governmental policy makers can obtain and analyze the continuous flow of commercial ratings data. They cannot, because of industry decisions made long ago, get universal, continuous information on more psychological aspects of viewing which could conceivably play a role in policy formation.

Diaries. The household tuning definition of viewing became unsatisfactory soon after the ARF recommended standards were issued. First, program demographic information became increasingly important to advertisers. Since passive meters could not provide this kind of data, a diary sample was added to the meter measurement system. In theory, the diary sample would provide continuous measurement of people watching TV, not just the set status. Nielsen was already working on its diary sample when the ARF report was published.

A second problem with the household tuning definition of viewing was that, as the number of multiset families increased, the notion of family viewing was less believable. The problem was more acute with radio (which Nielsen "audimeters" were originally designed to measure), since portability led to more out of home listening and undermined the notion of metering the a single household set. Personal diaries were developed to address the radio measurement problem. While television measurement still does not include out of home viewing, the household as the unit of measurement became less useful and more questionable. Meters had to be assigned to each set in the household.

Diaries also were employed for purely economic reasons. While a national sample of meters was maintained for the biggest clients -- the networks and the national advertisers -- stations in most local markets and their advertising clients have based their transactions on data from diary samples. Metering these markets has never been viewed as profitable by the measurement services, and the introduction of meters has been opposed by many broadcasters due to their extra cost and the belief that diary measurement may favor the older, established broadcast options, such as VHF network affiliates. Thus, metered measurement is reserved for the major television markets; a subset of them also are the ones who will receive people meters over the next few years.

(It is significant for FCC policy makers, who mainly face questions having to do with local markets, that most recent research and development in audience measurement technology has been directed at the problem of national audience measurement, as I review below.)

What phenomenon do diaries measure? Unlike passive set meters, diaries involve human discretion. People must decide if their behavior fits the one to be recorded, and then actually perform the work of recording. Neither task is easy. One is supposed to record continuously the time one spends "watching" television. If several people are "watching" at the same time, their individual behaviors are to be separately noted. The observations are to be made by the quarter hour. "Watching" five minutes out of the quarter hour is sufficient to be counted as viewing for the entire fifteen minutes. Households are asked to keep viewing diaries for seven days, one for each of the TV sets in the household. Having participated in a diary panel before, I can report that it is an onerous task, even for a household of only two people and two television sets.

A basic decision to be made in keeping a diary is when "watching" occurs. Nielsen and Arbitron leave this decision to the diary keeper to make -- they offer no rules or guidelines for making the judgment. This indeterminacy allows for considerable leeway, and one can only hypothesize about the rules followed in most households. One might expect the diary keeper to adopt a simplifying heuristic rather than to make ad hoc watching judgments whenever the set is on. One likely heuristic is to record oneself and others as "watching" if in the room with a set that is turned on. That rule is easy to follow, compared to one which would require judgments about attention to the screen. The "in the room" rule would result in overcounting of inattentive people, but would be more likely to count those with an opportunity for exposure. (People outside the room, but paying attention to the audio portion of a program, would be missed).

If these surmises make sense, diary measurement of television viewing would largely resemble individual tuning measurement, rather than focusing on the psychological aspects of viewing we discussed above. Another facet of diary keeping supports this interpretation. It is widely believed that diary keepers do not continually record viewing coincident with the behavior. Instead, much viewing is believed to be recalled at some later point. Assuming that this approach is common, it seems likely that the diary keeper could make an effort to recall bodies in the room during certain viewing periods, but is unlikely to be able to recall on a quarter-hour basis the attentiveness of various people, or even oneself. Finally, appeals to households to participate in a diary panel sometimes offer the appeal to "vote" for one's favorite programs, but this appeal does not map well onto the behavior required of the diary keeper. Diary keeping is more like filling out an expense report than casting a ballot. One is supposed to record precisely what one and others in the household did, regardless of the programming one would like to support. All of these factors militate toward diary measurement being like set tuning, since aspects of psychological involvement are too hard to keep track of accurately.

For the purpose of buying and selling time, this operational definition of watching may be perfectly adequate (although advertisers are growing increasingly restive with the concept of "opportunity" to speak to audiences). There are undoubtedly policy questions for which this approach is suitable. But there are other policy issues for which the "bodies in the room" definition would be unsatisfactory. Presumably this is why the BBC has run its ongoing program of audience satisfaction measurement for many years. The FCC is unlikely to address questions of audience satisfaction in the same way, but the point is that, even if an activist FCC wished to do so, it has no say in the design or operation of the universal, continuous measurement system that could provide such information.

People Meters. People meters, though certainly revolutionary in some respects, are a conceptual extension of diary measurement. One can view the act of measuring television watching with a people meter as using an electronic diary. Viewers are supposed to "punch in" when watching television, but the definition of watching is again left to the viewer. Using the same rationale noted in the section above, it is reasonable to expect individuals in people meter households to adopt a simplifying heuristic to decide when to punch in and out. "In the room" is a probable decision criterion. The finer discriminations made possible by the meter make the respondent's task more complex, however. Instead of recording in quarter-hour blocks, the people meter records by the minute, and a viewer's presence or absence during that short a span of time is a significant matter, particularly to advertisers in search of the grail of commercial, rather than program audience measurement. Thus, respondents in the people meter sample are called upon to be more accurate than their counterparts in diary panels, even though their button-pushing task is easier than filling in a diary. It is not clear how well respondents fulfill their assignment, particularly when it comes to short absences from the room. It is unlikely, however, that those using people meters take psychological criteria into account when deciding whether to log in or not.

If the concept of viewing is not so different in diary and people meter systems, the sample size and data delivery components of the Nielsen people meter system (the survivor at this juncture) are significant departures from the old set tuning and diary panels. The 4000 home Nielsen panel is over twice the size of the old NTI panel. Overnight data on individual viewership in all "in tab" homes is now available over night, with over 30 demographic categories summarized in standard reports. Clients access these data on line, and have the ability to customize analyses. The detail and volume of the data, and the speed with which it is delivered, present unprecedented challenges to users. Further, the sample size lends legitimacy to the idea that the national television "mass audience" is a dying concept. While even a sample of 4000 homes does not go far enough to appreciably improve measurement precision for low rated national viewing options (such as cable networks), the decision by Nielsen and AGB to employ samples in the 4000-5000 range is an important symbol, backed by an enormous financial commitment, that national television audience measurement no longer consists of network measurement and the standard demographics. One simply doesn't need the precision provided by samples that large to produce viewing estimates for women 18-49 in the Cosby audience.

There has been much heated debate over whether people meters should be introduced, and much discussion about how people meter systems should operate. Competition between Nielsen and AGB was the driving force in people meter development and the industry changeover. Some major clients, notably NBC, strongly opposed the people meter concept from the beginning. (NBC even threatened at one point to help develop an alternative rating service). As a result of deadlines and specifications set by the competitors, and pressures and decisions by the clients, we now have a new way to measure national television audiences. I needn't point out that, again, governmental policy makers were not players in any of deliberations that led to this transition. It could be argued that they needn't and couldn't have been involved, but policy makers certainly will have to live with the product if and when they base policy decisions on national audience data.

The Future. A number of observers argue now that people meters are but a transitory step in ratings methodology. Nielsen's stated objective is to develop and implement totally passive people meters. Various forms of technology for judging the presence of bodies in the room with the television have been investigated by various companies. R.D. Percy recently combined a people meter technology with a heat sensing device to check for people who have not punched in or out. Nielsen has experimented with passive measurement devices for decades, including couch meters, sonar, and metal detectors (which would require people to wear a badge of some sort).

On the assumption that some passive device will be introduced, the implementation will once again remove human discretion from the measurement process, and psychological aspects of viewing, however minor in their significance in diary and people meter measurement, will be completely eliminated. This may, like the original passive meter, make sense for industry purposes. A number of important industry actors, including Gale Metzger, who has designed and run numerous methodological studies for CONTAM and COLTRAM, argue that it is more important to get reliable measures of one aspect of the viewing phenomenon than to get unreliable measures of psychological ephemera. This argument runs directly counter to claims of academics, such as Jack McLeod at the University of Wisconsin, that attention is more important than mere "exposure" in explaining effects of viewing. It also runs counter to the numerous claims of "uses and gratifications" researchers, about the importance of motivations for exposure in explaining effects. To the extent that these researchers are correct and to the extent that policy decisions concern matters of effect (such as the "channeling" issue discussed below), the removal of human discretion from viewing measurement and the focus on presence in the room instead of on attention or involvement will deprive policy makers of information that could make a difference in their actions. Like other developments in ratings methodology, the policy-making user community will have no say about whether and when individual passive measurement is implemented.

Other future developments in commercial measurement of audiences may have a profound impact on the use of ratings data for policy-making. First, both Nielsen and Arbitron are developing and marketing national "single source" measurement services. "Single source" is a malleable term used to describe databases that contain information on audience exposure to marketing stimuli along with information on the audience's purchasing patterns. Originally used to describe syndicated self-report studies such as those of Simmons Market Research Bureau and Mediamark Research Incorporated, the term now is applied to measurement services that electronically monitor television advertising exposure, in-store promotion and coupon use, and product purchasing.

An example service is that provided by Information Resources Incorporated. In ten small cable markets, IRI measures television exposure (with passive set meters) and product purchasing (with in-store scanners). Panel households are recruited in grocery stores; they agree to carry an identification card to the store (grocery and drug stores which have been equipped with appropriate scanners) when they shop. Promotional activity -- e.g. special displays, cents off deals -- is recorded in the stores' computers. Purchases, identified by the identification cards scanned along with merchandise at the time of purchase, are linked with promotional activity and with television exposure, measured with one set meter per household. While much of IRI's research business involves promotion, new product, or advertising tests within markets, the company also has sought to pool its data over markets and sell a "nationally projectible" measurement of the link between television exposure and purchasing.

There are two more credible variations on the national single source theme. First, Arbitron teamed with SAMI-Burke several years ago to create a prototype single source system in Denver. This system paired a people meter procedure for measuring television exposure with a household "scanner wand" to measure purchases. The scanner wand is used to record purchases when they are brought into the household (by passing it over the bar code on each item). Since the Denver test, Arbitron and SAMI-Burke have gone separate ways, but each now claims to be developing a national system based on the prototype. Meanwhile, Nielsen has two systems in the planning stage: one based on the IRI identification card method, the other based on the scanner wand technology, and both with passive set meters.

The ID card method for measuring purchases is tied closely to the store as the foundation of the measurement system. Stores are equipped with the appropriate scanning and computing technologies by the research firm. Getting people to carry cards to the store in exchange for premiums is a relatively inexpensive addition to an expensive system for tracking in-store manipulations of display, pricing, shelving, and the like. Household panelists are actually recruited in stores in the IRI markets, while Nielsen proposes to sample households within store-based geographic units. The IRI approach bears no resemblance to a probability sampling method. The Nielsen sample is supposed to be based on probability methods, but, being tied to stores, will not be a probability sample of the populations within the selected markets.

The scanner wand liberates samplers from the geographic constraint, and permits measurement of products not sold in grocery or drug stores equipped with the right scanning technology. Obviously, the wand requires more effort by panelists than does carrying a card to the store, so fatigue, compliance and panel attrition are real concerns. This concern is heightened, when one considers the Arbitron--SAMI-Burke method, by the use of people meters for television exposure measurement, which also require considerable human effort. Of course, passive set meters get no person-level data of any kind, returning to the days before even diaries were adopted.

Single source plans and possibilities are interesting in their own right, but their significance for the future of ratings data is how they are perceived and evaluated by major clients, and how the investment in them by research firms affects their efforts in "regular" audience measurement.

We have a system of national and local audience measurement because of advertising. Networks and local stations have collected billions of dollars over the years because certain manufacturers believed that advertising on television was in some way related to sales of their products. These beliefs are being re-examined, challenged, and sometimes discarded because of single source data. So-called "weight tests" in IRI markets -- studies which compare purchasing in households who receive different doses of advertising for the same product -- have failed in most cases to turn up a relationship between advertising expenditure and purchasing. On the other hand, various sorts of promotion tests appear to have persuaded many clients of the value of coupons, displays, and cents off deals. Experience with single source data, however flawed that information is (and there are some notable design and conceptual flaws), has contributed to the shift in the industry-wide advertising/promotion spending ratio from 60/40 to 35/65 in the past decade.

Advertisers, particularly in the day of mergers, takeovers, buyouts and so forth, want short-term results. They also want to cut costs. Some of them are already querying Nielsen about why they should pay for two national systems of measuring television audiences -- one strictly for television, and the other for the television exposure/purchasing relationship. Niceties of sample representativeness, measurement error, and the like may be less impressive to them than the cost of supporting both systems, particularly when they are really most interested in the purchasing data. Thus, even though "national" single source systems do not and probably will not resemble national area probability samples, they may look "national" enough for some major advertisers, and even though the viewing data may be only set tuning, some clients may be willing to live with it.

AGB is reported to have lost some $100 million in its bid to compete with Nielsen. Dun and Bradstreet has certainly put an amount close to that into the Nielsen people meter system. The Nielsen single source plans, however, appear to dwarf that investment, with two "national" household panels of over 10,000 each, the scanning investment, automated commercial monitoring and other features. Even Dun and Bradstreet has limits on its ability to fund new ventures. Even Control Data may direct resources out of local market TV and radio measurement to support single source measurement. What will happen to "regular" audience measurement in the future, if the single source plans come to fruition and if major clients become less and less enamored of television advertising? One can imagine the marketing people in the measurement services figuring out ways to make 30 markets worth of TV data somehow "nationally representative." One can imagine clients, who fought hard for person-level TV data, now being willing to settle for household set meters, as demographic segmentation is discarded in favor of "behavioral" segmentation based on previous purchases.

The scenario is admittedly speculative, but not too far-fetched to suggest that syndicated audience data, used by clients and non-clients alike for many purposes (but designed only for audience commodity transactions) may become obsolete for its major purpose. If that happens, or if the designs of audience measurement systems are merely changed to accomodate new interests such as product purchasing, data users who are not party to the alterations will have to adapt their interests to the new design, or they will be out of luck. The scenario, to the extent that it has validity, applies to local market measurement as well, and even to measurement of other media audiences.

"Single source" is roundly criticized by representatives of radio and print interests because of its focus on television. The fear seems to be that advertisers will come to think of "media" as television. While there are many manufacturers for whom single source data, as presently configured, is not useful (the durable goods and business-to-business trade come readily to mind), users of broadcast ratings data depend heavily on the major packaged goods companies for the present design of the measurement system. These clients, for whom single source data is a most attractive lure, could help to shape audience data in ways that other media and policy makers find detrimental. At the very least, these speculations highlight the vulnerability of policy making based on audience data to changes in the industry for which the data were designed.

How good are ratings data now?

Thus far I have pointed out some of the consequences for policy makers of using ratings data over whose design they have no control. I noted that commercial audience measurement has stressed reliable reporting of behavior over psychological components of viewing. The notion of "tuning" has been emphasized over the idea of "watching," even when people have been called upon to play a role in the measurement. There are, however, some policy questions for which "watching," with all of its subjective connotations, may be a better definition of the phenomenon. Further, most recent research and development in the ratings business has focused on national measurement, and on methods for measuring phenomena other than viewing (i.e. purchasing). The industry is not at fault for having the wrong definition or investing its money in the wrong place. Rather, policy making based on someone else's operational definition may be misdirected, or, as in the single source scenario, vulnerable to changes in the system for which the data were designed, or to lapses in data quality because the information useful to the policy maker does not have a high industry priority.

I now want to consider the quality of ratings data for its stated purpose, and particularly focus on some contrasts between these data and those produced by ongoing federal surveys which form the basis of policy-making in many government agencies.

Sampling Error. One important characteristic of data quality is the level of precision in estimates permitted by the sample size and sampling method. Assuming a probability sampling method, the larger the sample size the smaller the sampling error, though a point of diminishing returns is reached as the sample gets larger (the sample size must be quadrupled in order to reduce sampling error by 50 per cent). An important consideration in determining needed sample size is the nature of subgroup analyses which may be performed. For example, if one wants to make estimates of ratings for particular demographic groups, or for individuals in particular geographic regions, the overall sample size must be large enough to pull in a sufficient number of such persons. This consideration becomes more important still for audience estimates for programs with low ratings.

Perhaps the most important dimension for the research firm in setting sample size is cost. The firm must be able to recoup the investment of installing meters or collecting diaries, and must take into account the amount clients are willing to pay for precision. The cost issue is particularly significant in local market measurement. The Madow Committee noted (p.28) that the rating services may lose money measuring some local markets, and an NBC executive recently told me that he thinks some local market measurement is subsidized by profits from larger markets. If these beliefs are correct, the cost environment of local market measurement militates against large sample sizes on a continuous basis, though they are necessary for precise subgroup estimates. (The cost environment also militates against more expenditure for other error reduction efforts, as we will see below).

The Madow Committee noted the "Catch 22" nature of the cost problem in local market measurement. On the one hand, measurement is less precise and less frequent; on the other hand, the amount of money to be made does not motivate more investment in better measurement:

> "There are many points where we feel [the rating services] should report more, study more, and do different things -- but we cannot claim that the market to which the rating services sell will pay the price..."
>
> Madow Committee p.12.

There are some 150 television markets that will continue to rely on infrequent diary measurement, as passive people meters, single source systems, and other electronic measurement systems, utilizing larger samples, are introduced in the larger markets. Significantly, many policy issues facing the FCC, tied to cable regulation or station practices, are local matters, and decisions must rely on data collection mechanisms that have been deemed outdated for some industry uses, but which persist because of cost considerations.

A more general point to be made concerning sampling error is the fact that ratings services do not perform strict probability sampling. Rather, when they find a predesignated household that refuses to participate in the panel, they substitute another "matched" household. Since the probability of a substituted household coming into the sample is unknown, the sample is not of the probability variety. Strictly speaking, then, rating services cannot calculate standard errors. They report standard error tables nevertheless, with the caveat that the reporting is not strictly appropriate.

Nonresponse Error. "Cooperation" rates -- the measure of what percentage of households sampled for ratings observations actually participate -- hovers between 40 and 50 per cent for national measurement and, in the aggregate, for all local markets. The last decade has seen a slow decline in cooperation rates in local markets (NAB, 1978, 1984, 1987), from just over 50 per cent to just below. The larger markets have been consistently lower. Similarly, the Nielsen people meter cooperation rate has ranged from the low 50s early in the service to below 50 now.

The impact of nonresponse on estimates of viewing is not easy to estimate or even study. There is a suspicion that those who do not agree to participate in ratings measurement may have different viewing patterns than those who do agree to participate. Using a telephone coincidental as the benchmark, a COLTRAM (1976) investigation found a small difference in the viewing of responders and non-responders, with responders doing more viewing. Coincidentals are useful techniques, but they cannot cover viewing patterns over a long time frame, nor viewing in certain dayparts. Thus, the closeness of the results of the COLTRAM research should not be viewed as definitive. Compared to governmental and academic surveys that average between 65 and 95 per cent response, the commercial rating services' response rates are very low. Many government agencies base policy on surveys that consistently achieve response rates over 90 per cent, whereas those faced with broadcast policy must be satisfied with a 50 per cent standard, to which most commercial users have become inured. This situation is another example of broadcast policy makers having to settle for what the industry has decided is all right, regardless of the importance of the data to the decision at hand.

Response Error. Here are a few examples of work on response error involving ratings data. The COLTRAM study mentioned above, in addition to other work suggests that response error may be a bigger problem than nonresponse error, at least where diaries are concerned. The COLTRAM research found that the diary underreports household and persons viewing, compared again to a coincidental benchmark. Similarly, the CAMS study, cited by Beville (1985), found that cable viewing was substantially underreported compared to a coincidental. Methodological work is still going on to investigate various error components of people meter measurement. Coincidentals to date point to children's measurement as a tough problem, as button pushing seems to be less than optimal for that age group.

Other Issues. This necessarily brief overview of some data quality matters concludes with some remarks about the system of methodological research and oversight for ratings data. The situation is clearly much better than when the Madow Committee examined ratings methodology. The many industry sponsored studies have given insight into the shortcomings and positive aspects of current methods. Perhaps more important, an oversight system and a set of procedural and disclosure standards under the auspices of the Electronic Media Rating Council accredits and audits services which desire industry approval. The auditing process insures against fraud (for those services that apply), and the disclosure standards are impressive; they are much more extensive than those of the American Association for Public Opinion Research, and the Council of American Survey Research Organizations, for example. Still, the EMRC system is hampered by the fact that it is voluntary. By contrast, all federal surveys must meet the strict standards of the entity performing the work (e.g. Census), the agency sponsor (e.g. the National Center for Health Statistics), and the Office of Management and Budget.

The current CONTAM evaluation of the Nielsen people meter system reveals both the value of industry sponsored methodological efforts, and the constraints surrounding them. The CONTAM work, being supervised and carried out by Statistical Research Incorporated, is inquiring into aspects of the Nielsen system ranging from how panelists were recruited, to the engineering of the meters, to panel management and response error correction. The reputation of the firm carrying out the evaluation is impeccable, and I suspect that much will be learned in the investigation. Still, there are roadblocks. For Nielsen, the study is a threat to their proprietary interest in the people meter system, and an adversarial atmosphere can develop between the evaluating agency and the measurement service. For example, SRI proposes to interview former people meter panelists concerning their experience. Nielsen has reluctantly agreed to the study, but has objected to giving away to competitors their hard won experience (since the CONTAM data will be in the public domain). Further, Nielsen has expressed some concerns about the use by the networks of the data from the interviews with former panelists. The fear is probably justified that, if problems emerge in the interviews, the networks, two of whom objected to the whole idea of people meters, will waste no time in hammering Nielsen's performance.

The atmosphere for the CONTAM study, then, is very different from methodological work carried out in an academic or governmental setting. The commercial basis for research in some ways discourages cooperation with outside methodological studies, and inhibits internal research, except for what the Madow Committee termed sales research and "crises" management. One of the surprising and somewhat disappointing facts about commercial research companies is that they rarely have continuing methodological research programs. Such efforts do not solve the immediate problems faced by the firms, and researchers are so busy trying to get the day's work done that conducting long range work is not a realistic possibility within their current structure. The fact that commercial research firms are primarily reactive in their methodological research (responding to clients' complaints or unanticipated problems) limits the scope and quality of the work, and hampers innovation. The fact that clients, who should apply pressure on research firms to get better quality data, are often short-sighted and niggardly further stultifies methodological work.

The lesson from all of these observations is that, while ratings research is much better today than it was in 1960, the quality of the data still is substantially below that produced for federal agencies by the federal data collection services or by the academic-based survey organizations. Second, while there is oversight within the industry, there is considerable leeway for the measurement service to avoid scrutiny and to avoid investing in methodological research. The Madow Committee recommended a centralized methodological research office for the industry. Instead, CONTAM, COLTRAM and the like were formed, and their efforts have been worthwhile. But their agendas are not set dispassionately, or with a long view. Further, their notions of "good data" are shaped by what the industry has come to expect to pay for research. The measurement services themselves do not have long term programs of methodological research; nor do they share the knowledge from ad hoc studies with the industry (except selectively in sales presentations). These constraints on methodological improvement may make sense for the industry, but do not necessarily fit well with the needs of a government agency for policy making.

How do policy makers use ratings data?

Finally, I want to look at a couple examples of use of ratings data by policy makers to point up some of the difficulties they face in dealing with numbers they had no hand in producing.

Consider the issue of "significant viewership" which first surfaced in connection with "must carry" rulings in local markets. Television stations within a 50 mile radius of the headend of a cable system had to show a minimum amount of viewership in order to fall under "must carry" guidelines. Webster (1988) has noted that the significant viewership concept has retained some currency even after the must carry issue was taken to court. He points out that significant viewership by county is calculated by the FCC, with some counties being represented by as few as 10 diaries to form the estimate. Such an estimate on its face has problems of reliability and validity, given what we know about the sampling and nonsampling errors to which ratings data are subject. Another example is policy making concerning children. Given what we know about the accuracy of estimates of children's viewing, it would be ill advised to make fine judgments based on ratings data. Further, "channeling" is a method which seeks to prevent children from being exposed to undesirable content. The policy assumes that some sorts of effects may result if children are exposed. In such cases, a psychologically-based definition of exposure may be more appropriate than a "tuning" measure, since attention to the tube is a likely mediator of such effects.

A major issue is the confidence placed in estimates of significant viewership and children's exposure to TV between 12 a.m. and 6 a.m., and the price of a wrong judgment in the policy question at issue (e.g. the channeling controversy). One approach for the policy maker is to use ratings data as a supplementary source of information which buttresses an argument based on other rationales. Another approach is to use ratings data as a central argument, but not for decisions that are very important. However the data are employed, the policy maker should follow the injunction of the President's Commission on Federal Statistics (1971):

> "Most policy decisions in the real world can be made, quite sensibly, by relying on less than perfect data, provided the users understand the dimensions of accuracy in the data they use ... Without such yardsticks, users may rely excessively on statistics, presuming an accuracy greater than exists." (pp.31-32)

The Commission goes on to state that many agencies have only the crudest understanding of the nature and degree of error in the data they use. They also noted that secondary users of data "must exercise great care in assessing the validity and reliability of the product. They should recognize that peculiarities may exist in definitions and error structure, and they should be responsible for inquiring into these matters." (p.55).

We have reviewed some "peculiarities" in the creation of broadcast audience ratings, and discussed some possible implications of them for the ability of policy makers to use the data in decisions. My particular aim has been to point out the price paid by policy makers when they have no role in the design of the data they use. The price of a custom study is easy to recognize. It is not always easy, in the pressured environment of the policy maker, to see the price of being a secondary data user. It is easy and tempting to treat the data as if they were collected with one's own aims in mind.

I argue that policy makers who rely upon ratings data should be aware of the forces that have and will shape the nature of the data. To the extent possible, they should encourage and participate in industry efforts to investigate and improve data quality. Where significant, continuing information needs are identified, policy makers should seek to develop their own methodologically sound data sources. At the very least, they should not seek to formulate policy mainly on the basis of data over which they have no control.

REFERENCES

Advertising Research Foundation (1954). "Recommended Standards for Radio and Television Audience Size Measurements."

Beville, H.M., (1985). Audience Ratings: Radio, Television, Cable. Lawrence Erlbaum.

COLTRAM (1976). "How Good is the Television Diary Technique?"

Madow Committee (1961). "Evaluation of Statistical Methods Used in Obtaining Broadcast Ratings," Report of the Committee on Interstate and Foreign Commerce.

NAB (1978, 1984, 1987). "The NAB Local Television Audience Trend Analysis." National Association of Broadcasters.

The President's Commission on Federal Statistics, 1971. "The Bureaucratic Pathology." U.S. Government Printing Office.

Spencer, B., and L. Moses, (1985). "Needed Data Quality for an Ambiguous Decision Problem." CERA Working Paper, Stanford University.

Webster, J. (1988). "The Role of Audience Ratings in Communications Policy." Northwestern University. Institute for Modern Communications. Working Paper.

OPEN INTERCONNECTION
AND THE ECONOMICS OF NETWORKS:
An Analysis and Critique of Current Policy.

Milton Mueller
4405 Osage Ave.
Philadelphia PA 19104
215-243-0160 (h)
215-898-6017 (o)

University of Pennsylvania
Annenberg School of
Communications
Dr. Carolyn Marvin,
dissertation advisor

Introduction.

This paper investigates one of the fundamental assumptions underlying contemporary telecommunications policy. The assumption in question is what I will call the principle of open interconnection, and includes a cluster of beliefs about the interconnection of competing carriers and the effects of the same on telecommunications markets. In essence, this view holds that the once-monopolistic public telecommunications network can be made as competitive as the neighborhood grocery store if all carriers are required to interconnect with each other according to uniform technical standards and nondiscriminatory tariffs. Interconnection is believed to open telecommunications markets to free competition by ensuring that neither compatibility standards nor network externalities can work to reinforce the dominance of an established network. A carrier's refusal to interconnect its network with other systems is seen as an inherently anticompetive exercise of monopoly power.

It is no exaggeration to say that interconnection policy is the keystone of the contemporary response to the problem of monopoly.[1] With roots in older antitrust and common carrier principles, the ideal of open interconnection has been applied with increasing consistency from the days of the MCI decision in 1969 to the Equal Access and ONA proposals of the 1980s. Yet the principle itself and the economic assumptions underlying it have never been subjected to much critical scrutiny.

From a historical standpoint, open interconnection is a major policy innovation. In the "bad old days" of telephone competition in the early 1900s, competition meant that telephone subscribers were fragmented into two or more unconnected systems. Those who wanted access to everyone had to buy duplicate subscriptions. By 1910 the country was faced with what appeared to be a tough choice. Competition, with its innovation and price constraints, could only be purchased at the price of fragmenting subscribers. Monopoly could bring universal interconnection of telephone users, but eliminated price competition and the technological dynamism of the market. The perception that there was an inherent, irreconcilable conflict between universal service[2] and competition was the primary motive behind the choice of regulated monopoly as the institutional framework of telecommunications. Today's open interconnection policy, however, appears to sidestep the whole dilemma. It promises competition without fragmentation; universal access without monopoly or rate regulation. It almost sounds too good to be true.

It may be. This paper challenges much of the conventional wisdom concerning interconnection, competition and monopoly. Using historical cases, it demonstrates that interconnection tends to make networks complements rather than competitors, and can easily preempt network competition rather than promote it. Conversely, the noninterconnection of alternate networks can be both procompetitive and socially beneficial in certain ways. Much of the theory supporting open interconnection, in fact, is based on an incorrect reading of telephone history.

The paper reviews the theoretical literature supporting open interconnection and shows how it fails to recognize a crucial distinction between the effects of interconnecting networks and the effects of mandating compatibility in equipment. Networks have a special economic trait, which I call nonhomogeneity, that makes the effects of mandated interconnection more problematical than the standardization of equipment interfaces.

The paper is cast as a critique of some of the assumptions underlying current policy. It does not attempt to propose a full-fledged alternative policy or theory. It is intended to raise questions rather than to answer them; to stimulate debate over issues that have not been adequately considered rather than to define a new direction.

I

Theoretical Foundations of Open Interconnection

Two distinct but closely related bodies of theory provide the foundations of the current approach to interconnection. One group focuses on the competitive effects of standardization and compatibility in system-selling industries. The other group analyzes the competitive consequences of the "network externality," a name coined by economists to describe products that become more valuable as more people use them.

Competition and compatibility.

The literature on the economics of standardization sprang into existence in the mid-1970s in connection with the IBM and AT&T antitrust cases, and has been steadily expanding ever since. It ranges over subjects as diverse as bank cards, railroad gauges and the effects of standardization on product quality. I will be concerned here only with those works that deal with the strategic uses of compatibility in a competitive environment.

Much of this theory is a rather straightforward application of the concepts of tying and vertical foreclosure in antitrust law. According to this view, a seller with market power in one commodity can require its users to purchase a second good by tying or bundling it with the first, or by refusing to permit consumers to use competitively-supplied complementary goods. Both activities are believed to injure buyers by restricting competition in the market for the second good, and injure other sellers by acting as a barrier to entry.

In the mid-70s, economists discovered that technical compatibility relations could be used as tying mechanisms and barriers to entry. By deliberately designing products so they would not work with the components or systems of other manufacturers, a producer could lock buyers into his product line and shut other producers out of the market. They also observed that businesses with large market shares exerted disproportionate influence over the process by which industrywide standards were set. A successful standard-setter thus occupies the "dual role of rival and rulemaker"[3] and is well-positioned to exploit compatibility as an anti-competitive weapon. Examples were provided by studies of the behavior of IBM in the 1970s[4] and of Kodak's dominance of the photography industry.[5]

This line of analysis lent itself to clear policy prescriptions. System markets dominated by a single producer with the power to unilaterally establish (and strategically manipulate) standards could be opened to free competition by setting uniform, public standards. Standardizing the interface between computer components, for example, permitted manufacturers who could not compete with IBM in the production of complete computer systems to enter the market for individual components. It also allowed consumers to "mix and match" the products of various manufacturers. The literature contains some recognition that competition between complete systems is possible without compatibility,[6] but puts overwhelming emphasis on the procompetitive effects of compatibility.

The same logic was readily applied to telecommunications. Indeed, the theory recognizes no qualitative distinction between telecommunications networks and computers, cameras or any other kind of equipment. The telephone network was viewed as a big, publicly extended "system" like a computer, with the local exchange, terminal equipment and long distance as its "components." AT&T's refusal during the '60s and '70s to permit the connection of rival long distance networks or terminal equipment was the economic and moral equivalent of IBM's use of incompatibility to thwart competition. It extended monopoly control from one market, the local exchange, to other, potentially competitive complementary goods.

Of course, economists recognized that compatibility _per se_ was not the issue in this case. The rival networks and terminal equipment were capable of working with the AT&T network, but the dominant carrier was unwilling to connect them. AT&T's behavior amounted to vertical foreclosure rather than tying.[7] The economic analysis identifying this behavior as "anticompetitive," however, was identical to that which had supported the standardization of computer component interfaces.

Network externalities.

A product is said to possess "positive consumption externalities" when one consumer's utility is enhanced by an increase in the number of other consumers of the same product.[8] The concept is particularly applicable to networks of communication, the value of which directly depends upon how many other people join the same network. This aspect of telephone service was recognized long before economists formalized the concept,[9] and was often used in conjunction with arguments for a monopoly telephone system.[10]

Because their value as a means of coordinating product design depends upon how many other producers use them, industry-wide compatibility standards are strongly characterized by network externalities. The literature on compatibility standards and network externalities thus overlaps to a significant degree.

Competition in markets with network externalities has special characteristics. Because much or all of the value of the product is derived from the fact that everyone else is using it, competition can become a battle for mutually exclusive control of all of the market rather than for a profitable share of the market.[11] For example, the manufacturers of competing AM stereo equipment seek to establish their AM stereo format as the de facto standard. Should one of them succeed, the equipment of the losing manufacturers will become obsolete because of its incompatibility with the dominant standard.

Network externalities give the competitive process a "bandwagon effect;" consumers will wait for a product or network to establish a clear lead over its competitors and then collectively flock to it, making its success self-reinforcing. Once dominance is attained, moreover, it can become self-perpetuating. The continued use of the QWERTY keyboard design despite the availability of more efficient layouts is a common example of what Farrell and Saloner call the "excess inertia" of an established standard.[12]

From a policy standpoint, the network externality concept appears to point clearly in the direction of universal interconnection. Dividing subscribers reduces the value of communications networks. The theory also suggests that large, established networks have inherent advantages that cannot be overcome without resort to compulsory interconnection with smaller rivals. Refusal to interconnect gives the larger network the full benefit of its superior size while imposing risks and costs on those who might choose to switch. Interconnection neutralizes the competitive advantage gained through the network externality and makes competition a matter of price differences.

The received economic wisdom regarding interconnection can be summarized as follows:

A) The competitive issues raised by network interconnection are no different than those raised by equipment compatibility. Both are treated within the framework of the tying and vertical foreclosure concepts, which emphasize the procompetitive effects of breaking up integrated systems into compatible-component markets.

B) Requiring a dominant carrier to interconnect with its rivals promotes competition by eliminating vertical foreclosure and overcoming the inertial advantage created by the network externality.

C) A dominant network provider will refuse to interconnect with a competitor except under legal compulsion. Refusal to interconnect can always be interpreted as anti-competitive, and has no socially redeeming features.

If we were writing a television commercial for the policy, we could claim that it opens "bottlenecks," prevents "lock-in," and stops "leverage."

II

History vs. Theory

Much of the prevailing theory about interconection has its roots in historical analysis. The conclusion that dominant systems or networks will not interconnect with smaller ones, and that the effects of this practice are anticompetitive, is supported by reference to AT&T's early history. The standard interpretation holds that AT&T's refusal to interconnect its long distance lines and local exchange facilities with the independents undermined the viability of the competition and eventually brought them to their knees.[13] The actual history of Bell's interconnection practices, however, tells a different story.

Bell's interconnection policy went through a gradual process of relaxation from 1902 to 1913. Table 1 sets out the various stages and shows how they are related to the status of independent competition. The Table and the accompanying statistics show that the relationship between interconnection and competition is almost the opposite of that posited by the conventional theory. The independents flourished under the most restrictive interconnection practices and declined as Bell policy was liberalized. Furthermore, much of the change was driven by market rather than political pressure.

From 1894 until about 1902, Bell exchanges would not connect with any independent, period. This policy was totally ineffective at containing independent growth, as is evident from a glance at the statistics. During those years 6,608 independent

TABLE 1:
INTERCONNECTION VS. COMPETITION, 1894-1913

A = Percentage of total U.S. telephones owned by independents not connected to Bell

B = Percentage of noncompeting independent telephones connected to Bell.

C = Percentage of total U.S. telephones owned by Bell.

DATE	BELL INTERCONNECTION POLICY	A	B	C
1894	No connections with any Independent exchanges. Refusal to sell Western Electric telephones to Independents.	5%	0	95%
1898		32%	0	68%
1902	Sublicensing: toll connections with noncompeting Independent exchanges. Bell demands exclusive connecting rights and use of WECo phones exclusively.	47%	12%	49%
1906	gradual relaxation of requirement that sublicensees use WECo phones	44%	27%	51%
1907	Vail OKs sale of WECo phones to all Independents and sublicensing without any use of WECo phones	37%	46%	49%
1909	(1908-1910) Long distance connections with some competing Independent exchanges	26%	79%	50%
1912	(Jan 5, 1912) "Vail Commitment." Competing exchanges acquired by Bell will preserve all preexisting long distance connections.	17%	na	52%
1913	Exclusive connecting contracts ruled illegal (Jan); Kingsbury Commitment (Dec): Bell provides toll connections to exchanges more than 50 miles apart.	15%	89%	55%

Data source: Bell Labs archives, Warren, NJ.

exchanges were established and the number of nonconnected independent subscribers grew to 970,000, about 47% of the total telephones in the U.S. The independents took root in the smaller towns and rural areas that had been neglected by Bell. As they gained momentum, however, they increasingly set up exchanges in cities already occupied by Bell. By 1898 32% of all Bell exchanges in cities with populations over 5,000 had competing independent exchanges operating alongside them; by 1902 the number had grown to 56 percent.[14] In negotiations with city governments, the upstarts often used their exclusive access to independent exchanges in surrounding areas as a compelling argument for a second franchise.

Interconnection as cooptation: the sublicensing policy.

After absolute noninterconnection failed to stem the rising tide of competition, Bell in 1902 turned to the policy of "sublicensing" independent exchanges.[15] In the sublicense contract, Bell agreed to interconnect its toll lines and exchanges to an independent local exchange, subject to three conditions: the independent had to be located in a community with no Bell exchange, it had to lease Western Electric-manufactured instruments, and it had to agree not to connect with the toll lines of Bell's competitors. In return, Bell promised not to set up a competing exchange.

Sublicensing was an important departure from its prior belief that Bell and Bell alone ought to control the nation's telephone network. In the internal debates over the new policy, its advocates pointed out that it was better to relax the interconnection policy and gain access to communities where there was no Bell exchange than to allow nonconnected and potentially competitive networks to develop.[16] In other words, sublicensing was a method of pre-empting exchange competition and undermining the long distance development of the independents, and was immediately recognized as such by the financial and ideological leaders of the independent movement.[17]

For the next eight years, sublicensing proved to be one of the most effective means of checking the spread of independent telephony. In 1903, the Bell company in the Ohio, Indiana and Illinois region had 114,000 subscribers of its own and was connected to 20,000 sublicensed subscribers. It was widely perceived as being on the verge of collapse in the face of the intense competition. A new manager, L.G. Richardson, took control in May and began to emphasize sublicensing. "We have a well-organized force of men in the field working on the sublicense proposition all_the_time," he wrote to national headquarters.[18] By January 1, 1908, the number of Central Union subscribers had grown to 188,000, a respectable increase--but the number of Central Union sublicensee subscribers had surged to 192,000.[19]

Connecting with independent exchanges took the wind out of the sails of independent long distance companies, which could offer exclusive connections to fewer and fewer locations. Significantly, however, Bell was forced to relax the conditions it attached to interconnection. Eventually it abandoned both the demand that Bell instruments be employed and even the stipulation that no other long distance company could terminate in the exchange.[20] Table 1 shows the dramatic increase in the percentage of noncompeting independents who agreed to connect with Bell as its policy was liberalized. Starting with only 12% in 1902, Bell by 1909 had won over 79% of the independents in communities with no Bell exchange.

The story so far conflicts with the conventional wisdom on several counts. First, the refusal of an established network to connect with its competitor does not necessarily make survival of the competitor impossible, even when the established network has a 20-year head start and dwarfs the newcomer in size. Second, a larger competitor may interconnect with smaller competitors without legal coercion if it fears that failure to bring them into its system will isolate it from significant markets and/or will provide the nucleus of a larger competitive system. In other words, smaller competitive networks do have appreciable bargaining power in their relations with larger networks, especially when the established network is unable to develop the market fully. Third, and most important, interconnection can be a powerful method of preempting competition between end-to-end systems. Networks have a natural incentive to enter the same territories and markets and compete when they are not connected, because the absence of interconnection forces them to build duplicative facilities to gain access. When they are connected, they tend to cooperate and divide territories and markets.

Equal Access precedent.

My next example is an "equal access" case that is directly comparable to our present policy, except that it took place 80 years ago, between 1908 and 1912.

Sublicensing notwithstanding, until 1906 the Bell system steadfastly refused to connect with independent exchanges in towns where there was also a competing Bell exchange. In response, the United States Telephone Company was organized in 1900 to serve as a long distance company connecting the many independent exchanges in Ohio. U.S. Telephone's toll contracts called for exclusive connecting rights with independent exchanges for 99 years. In this respect, they were no different from the Bell System's contracts with its licensee companies, which also established exclusive connecting rights with the parent company, except for the fact that U.S. Telephone usually did not acquire an ownership interest in the connecting company.

By 1908, U.S. Telephone was furnishing long distance service to several hundred independent exchanges reaching 325,000 telephones.[21] Because of its success, Bell was forced to reverse its policy of refusing to make toll connections with competing local exchanges. In late 1906, it began to actively compete with U.S. Telephone for the long distance business of independent exchanges directly competing with Bell System exchanges. In Richmond Indiana, for example, its local exchange had 109 subscribers and the independent Home Telephone Company had about 2,500. Bell proposed a connection agreement with the Richmond Home Co. that would give its long distance lines access to the independent exchange and vice-versa.[22]

Connecting with both Bell and Independent long distance lines clearly made all but the most ideologically anti-Bell local exchange managers happy;[23] they received connections to the subscribers of both systems. But just as clearly, it threatened the viability of the U.S. Telephone Co. In 1908 the independent long distance company sued four local exchanges in Ohio for breach of the connecting contract, and won an injunction from the Ohio Common Pleas Court. The case was appealed by the Bell interests to the Ohio Supreme Court, and went from there to the highest federal court in the state, the U.S. Circuit Court of Appeals.

The Court's decision, made in January of 1913, prohibited U.S. Telephone's exclusive toll connecting contracts as monopolistic and against public policy. Its reasoning will sound familiar to any contemporary advocate of the anti-tying view:

> The local company has tied up its long distance business. It cannot take general advantage of competition from time to time arising, no matter how advantageous to it or its patrons, and it cannot expand its own business beyond its then existing limits into competition with the long distance company, no matter how advisable such extension and competition might prove to be.[24]

The court held, in effect, that only equal access to the local exchange could preserve competition in long distance. The court's belief that exclusive tying contracts suppress, and that nondiscriminatory access to the exchange promotes, competition is still the prevailing view. In point of fact, however, striking down the exclusive toll connecting contracts was a major factor leading to the elimination of competition between Bell and the Independents in Ohio and Indiana. Once it was assured of access to independent local exchanges, the Bell System had no need to maintain duplicate exchanges in locations where the independent was stronger. In cities where Bell was stronger, the independent exchange could not benefit from being the exclusive avenue for making toll connections to other independent exchanges in the region. Frequently, they sold out and left a local monopoly to Bell. At the same time, the U.S. Telephone Co. was in no position to compete with Bell's extensive long distance lines without exclusive access to independent exchanges. The independent long distance company sold out soon after the decision.

The collapse of independent long distance competition at this time was followed by demands for open access to AT&T's Long Lines, and created the political pressures that led to the Kingsbury Commitment in December of 1913. Far from preserving or aiding independent toll competition, the Kingsbury Commitment was tantamount to its obituary. It signalled the independents' willingness to rely on the Bell system as the nation's backbone network and the end of their attempts to compete with it.

The Table 1 indicates that interconnection was more effective at eliminating competition than acquisition. The percentage of Bell-owned telephones only increased from 49% in 1902 to 55% in 1913, while the percentage of subscribers in noncompeting independent exchanges connected to Bell went from 12% to 89% in the same period.

Network competition and unequal access.

At both the local and long distance levels, interconnection quickly resulted in complementarity, not competition--in interconnected monopolies in separate territories, rather than in competitive rivalry conducted by overlapping systems. That the actual consequence of interconnection was to eliminate rather than promote competition is clear; the historical record leaves little room for doubt about this. What is less clear is why this should be so. The explanation is that the fundamental force driving telecommunications competition at that time was not differences in technology or costs, but differences in access; i.e., in who was connected to whom. The very thing that sustained multiple, rivalrous, overlapping telephone companies was the fact that none of them had universal, nondiscriminatory access to all the telephone users or exchanges in the area. Consequently, one telephone company could use its so-called "bottleneck" control of some exchanges to win subscribers and business in the territory of another telephone company. By the same token, its "bottleneck" status assured that anyone who wanted to compete with it had to build duplicate facilities in the area it controlled if it wanted to be able to reach subscribers there. Unequal access was the prime mover of competition.

Looking back on this experience, the tendency now is to emphasize the costs of this kind of competition--the need for businesses to take out duplicate subscriptions, the fragmentation of subscribers into unconnected networks, the unstable financial conditions brought on by competition.[25] There are, however, positive social benefits that should not be overlooked.

The experience of 1894-1913 shows that competing networks that are not interconnected have a powerful incentive to be as universal as possible. The network externality places a premium on universality and turns competition into a race to connect as many subscribers and locations as possible. Where the networks overlap, competition takes the form of attracting new subscribers with lower rates or better service, thus overcoming the alleged underproduction incentive created by the network externality. In this kind of competition, there is no potential for creamskimming; each network must strive to duplicate the facilities of its rival. The Bell policy of maintaining exchanges in unprofitable locations for the sake of universal service emerged during these years.[26] Had there been interconnection between Bell and its rivals, either network could have relied on other networks to service the unprofitable areas.

The U.S. achievement of universal telephone service must be attributed in large part to the rapid development that occurred during the competitive era,[27] and particularly to the structural characteristics of competition between nonconnected, end-to-end systems. Superior universality was AT&T's greatest advantage in the competitive battle. While the benefits of telephone competition in this period have been amply documented by other writers, they have not realized how many of those benefits were contingent upon the rival networks not being connected. Instead, following current doctrine, they tend to suggest that interconnection would have aided competition.[28]

III

Contemporary Applications

Does interconnection have the same effects on competition today? On the face of it, it would appear not to. The emergence of long distance competition and local exchange bypass has taken place alongside--some would say because of--open interconnection. On closer inspection, however, the pattern of complementarity emerges. Consider first the case of long distance competition.

At the time of the FCC's Specialized Common Carrier decision in 1970, the new private line carriers said that they intended to build their own local distribution facilities. Some preferred to assume end to end responsibility and had no interest in interconnection with Bell.[29] Others planned to rely on interconnection with AT&T's local facilities in some cases, but the record gives the impression that they viewed this as a temporary expedient.[30] Whether the alternative carriers could have survived without interconnection is a moot point, however, for the FCC made AT&T supply connecting facilities to them from the beginning. The new carriers were authorized to provide private line service alone, which involved connections with a few large customers and comprised only about 3% of the total market. This made the interconnection issue appear to be less important than the need for specialized services which, according to an abundance of testimony, AT&T was not adequately supplying.

MCI's surreptitious entry into switched MTS service in 1975 abruptly changed all that. By combining its own microwave network with local Bell switched services, MCI was able to avoid the separations and settlements process and offer discounted switched long distance service to a mass business market. Once its right to do so had been decisively upheld by the courts,[31] the policy of open interconnection of competing carriers was partially operative in the long distance market. It became fully operative following the divestiture and the implementation of equal access.

Clearly, the fledgling specialized carriers of the early '70s could have been a major force in the development of alternative local distribution facilities. Equal access interconnection, however, made it unnecessary for them to do so. As a result, today, ten years after Execunet, interexchange carriers rely on BOC local exchange facilities to carry 99.9% of their interexchange traffic and to reach 99.9999% of their customers.[32] Interconnection has produced complementarity rather than competition. It has allowed competitors to exploit the artificial gap between prices and costs in the long distance arena while leaving the job of terminating traffic to the Bell companies.

Open Network Architecture affords another example of the not-so-competitive effects of open interconnection.[33] The program is intended to unbundle basic network service elements in order to allow enhanced service providers equal access to the public through the BOCs' "bottleneck" local facilities. Yet the BOCs themselves openly view ONA as a way of discouraging the development of alternatives to their networks. Enhanced service providers have various technological alternatives to the local exchange, ranging from FM subcarriers to CD/ROM discs delivered through the mail.[34] By opening up their network to facilitate its use by enhanced service providers, the RBOCs aim to keep them on the public network, and to increase the utilization of their network.[35] The BOCs also realize that they have many valuable services to offer enhanced service providers, such as billing and joint marketing arrangements. As with sublicensing in the early 1900s, interconnecting suppliers tends to make them complementary rather than competitive.

Interconnection does prevent an established network from deriving a competitive advantage from its control of access to subscribers. But by doing so it largely abolishes competition in the production of access itself. Indeed, the existing regulatory process not only gives telecommunications companies the _right_ to interconnect, but also considerable influence over the terms and conditions of interconnection. The equal access and ONA programs amount to nothing less than a publicly administered reconfiguration of the local exchange network to suit the needs of large users. Why, given these conditions, would anyone bother to construct an alternative local exchange? Competition becomes a matter locating cost and rate disparities in discrete markets.

IV

Networks and Homogeneity

In the preceding section interconnection was analyzed largely in terms of historical cases. The next section attempts to give the analysis a theoretical grounding. In what follows, I will elaborate on the subject of the homogeneity of networks and show how it is central to interconnection policy.

Interconnection is normally discussed by economists as if it were a straightforward compatibility issue. In terms of the economic and competitive issues it raises, the interconnection of telecommunications networks is supposed to be no different in principle from the standardization of the interfaces between computer components, photography equipment, or telephone CPE. Farrell and Saloner, for example, use the equal access program as an example of "standardization."[36]

There is a difference between the two, however. The importance of that difference can be ascertained by contrasting the FCC's CPE registration program, which standardized the interface between terminal equipment and the AT&T network, with the interconnection of local and competing long distance networks brought about by the divestiture and the equal access program.

CPE compatibility did exactly what the economists said it was supposed to do. It abolished tying between AT&T's network facilities and its terminal equipment, thereby allowing hundreds of equipment manufacturers to compete fairly on the basis of price and design. The result was lower prices, greater product variety and an almost total replacement of regulation by market forces. The CPE registration program was also relatively easy to implement and enforce.

The attempt to create an analogous "standardized interface" between the local and long distance networks was anything but simple. It required ripping apart the telephone company, technically redesigning the whole local-long distance interface, and completely revising the flow of revenues in the telephone system. The process did not decisively replace monopoly with competition; it merely shifted the problem of monopoly from the long distance (interLATA) level to the local (LATA) level. Ironically, all of the same problems that led to the breakup of AT&T still exist in the LATA. Nor did divestiture replace regulation with market competition; most of the old system of utility regulation is still in place, while new forms of regulatory intervention have been added as a result. The local and long distance rates of AT&T and the BOCs are still regulated, the FCC has become heavily involved in the design and standardization of network interfaces, and severe line-of-business restrictions have been placed on the BOCs.

Why has standardizing the interface between competitive networks proved to be so recalcitrant compared to our experience with equipment standardization? Some blame the stubbornness of Judge Greene or the failure of politicians and the FCC to deliver on the implicit promise of deregulation after the divestiture. Others blame the general incoherence of the multilayered telecommunications policymaking process in the U.S. All of the above may be contributing factors, but the real causes are the unique economic characteristics of networks.

Networks as inherently nonhomogenous goods.

Competition as we normally define it requires that the products of rival producers be substitutes for each other. Enhanced substitutability accounts for much of the pro-competitive effects of standardization. Once full compatibility exists, an Epson printer is a substitute for an IBM printer. Likewise, once the interface between the network and terminal equipment is standardized, a Panasonic telephone is a substitute for an AT&T telephone. By enhancing substitution, standardization intensifies competition and expands the consumer's range of choices.

On the other hand, the MCI and Sprint networks are not full substitutes for the AT&T network, nor are local exchange bypass facilities full substitutes for the BOC network. By themselves, the alternative networks provide communications access to only a small portion of the people and locations served by the dominant carrier. In these cases the competitive alternatives are only partial, imperfect substitutes; in fact, their commercial viability relies very heavily on the access channels supplied by their rival.

Economists who have studied networks have failed to recognize both the uniqueness and the importance of the substitution issue. Katz and Shapiro's model of network competition is based on the assumption that networks are homogenous goods if they are of equal size.[37] This means that consumers will view any two networks as perfect substitutes as long as their gross number of subscribers is equal. When applied to real-time access networks, such an assumption is more than unrealistic; it obliterates a crucial feature of competition. Two telephone networks can be of equal size but serve completely different subscribers or locations. In this case the two networks are not perfect substitutes but entirely different products. Even when two equally-sized networks contain some duplicating subscribers or overlapping service areas they are still only partial substitutes.

The only way multiple networks can compete as homogenous products is for each of them to offer access to the exact same subscribers. Here we confront the central paradox of network economics. Only by interconnecting multiple telecommunications networks can we make it possible for them to offer a homogenous product. Yet interconnecting rival networks tends to make them complementary rather than duplicative, and thus imperfect substitutes for each other. This point requires some elaboration.

Assume, first, two competing, noninterconnected telephone networks, A and B. Because they are not connected, anyone who subscribes to B alone will not be accessible to the subscribers of A, and vice-versa. Only those subscribers who value universal access enough to pay extra will purchase subscriptions to both networks, and hence will be available to the customers of both networks. In order for the two networks to offer a homogenous product, 100% of all telephone subscribers would have to buy subscriptions to both A and B. That degree of duplication, however, is virtually impossible as long as the cost of a subscription is greater than zero, for two reasons. First, there will be low-usage subscribers on both networks who will not value universal access enough to pay for two subscriptions. Second, even if there was no such economic barrier, 100% duplication is self-negating. If all of the subscribers to A also subscribed to B, then any subscriber could cancel his subscription to one of the networks without losing access to anyone. In other words, 100% duplication would require each person on each network to pay twice for a subscription when they could already call all other telephone users through only one of the networks. Thus, duplication will never reach 100%; noninterconnected networks will always offer access to different subscribers and hence will be more or less imperfect substitutes.

Now suppose that A and B are interconnected. Will this give us homogeneity? It will certainly make the networks perfect substitutes in terms of the people and places accessed. But as we have seen, interconnecting rivalrous networks eliminates the incentive for them to duplicate each other. It allows network A to construct access facilties only to the places where it has a cost advantage over network B while relying on B to provide access to the rest. It also permits either network to refuse to compete in markets which they feel cannot sustain both businesses. Direct rivalry will be confined to a few choice markets and locations. Or, in cases where there is a universal monopoly beset by new entrants, competition will emerge in network segments that are undersupplied or overpriced, but the new entrants will still rely on the dominant carrier to supply access to all other segments. Once interconnected, networks behave more like complements than substitutes.

The nature of substitution in telecommunications can be clarified by the following representation (see Figure 1). This simple model represents the network as a matrix, with each network terminus assigned a row and column. The letters A - n represent subscribers or locations. The values in each cell are arbitrary values chosen to represent the amount of traffic or information transmitted between the two points. The object of representing a network this way is to show how tricky it can be to define the "components" of a network. Each individual cell and every possible combination of cells in the matrix represents _a separate region within which substitution can take place_. Each cell (and every combination thereof) differs from every other in economically important ways.

FIGURE 1

Components of a network connecting A - n points.

	A											
B	1	B										
C	1	2	C									
D	1	3	1	D								
E	2	2	3	1	E							
F	6	8	5	2	1	F						
G	1	4	1	2	5	9	G					
H	2	12	7	6	5	3	1	H				
I	4	18	9	8	4	4	3	3	I			
J	8	2	8	20	11	2	1	5	3	J		
K	16	7	13	50	9	4	6	9	12	2	K	
n												n

Suppose that B represents a corporation's office in Los Angeles, H represents a branch office of the same corporation in San Francisco, and the business is comparing the costs of constructing a private network with the costs of using the public switched network. In this sense, the private system "competes" with the public switched exchange for the market represented by the shaded cell in Figure 1. But the type of substitution going on here is not the same as making a choice between two brands of dog food or two types of computer components. The "market" in which the buyer shops is not something generic and homogenous like "local exchange" or "long distance" service. It is very specific, and unique to that particular buyer. In terms of both required circuit and switching capacity and distance, the cost of linking those two points will be different from the costs associated with linking any other two points on the map. If any other points are added, e.g. if the business considers adding G to its private network, the whole cost calculus changes. Telecommunications networks are not a single product but bundles of many complementary products, none of which are identical.

Considered in isolation, each cell represents a potentially homogenous good. If multiple carriers vie for the traffic represented by each individual cell, and none of them combine the operations required to serve that cell with any other cell, perhaps something similar to what economists define as perfect competition can take place. But merely to raise this as a possibility is to demonstrate its impossibility. There will always be economies of scale involved in aggregating and concentrating traffic from more than one cell, especially when low density cells are involved. The most likely result of open interconnection is a patchwork of nonhomogenous networks with competitive substitution confined to a few high-density cells. The networks involved will be mostly complements, and only partly competitive substitutes.

IV

The Policy Dilemmas of Nonhomogeneity.

With the foregoing in mind, we can begin to analyze why the interconnection of competitive telecommunications networks poses more problems than mandating compatibility in equipment markets.

Current economic doctrine defines the components of a system as complementary goods, and holds that tying or anticompetitive pricing can be eliminated by unbundling the components via standardized interfaces. Unbundling lowers barriers to entry and intensifies competition by allowing substitution to occur in each component of the system. Interconnection has a similar effect on telecommunications networks. It allows competitive substitution to occur along specific routes or lines rather than between complete systems. But that is where the similarity ends. The "components" of a network are not identical, mass produced objects like computer printers or dog food. They are hundreds or even millions of communication links whose economic and physical characteristics are all different. The components of a communications network cannot be defined in generic terms. There are as many different components as there are combinations of users.

The local-long distance interface provides an excellent example of how this nonhomogeneity throws a monkey wrench into attempts to apply compatible-component thinking to networks. Economists have managed to convince themselves that local and long distance are separate components of the network.[38] No hard and fast boundary between the two can be defined, however; the distinction is purely a matter of degree. The exchange business has always included the provision of toll service to suburbs and nearby cities. Historically, the Bell operating companies were responsible for regional long distance development. Sometimes the people connected by a network will be close together and sometimes they will be far apart, but there is no objective basis for dividing service components.[39]

In attempting to draw component boundaries where none exist, the legal and regulatory system came up with Local Access and Transport Areas (LATAs). Predictably, LATAs are utterly arbitrary, being limited to one metropolitan area in some cases and encompassing entire states in others. As an attempt to distinguish local from toll business, they are a total failure; close to 40% of all toll calls are made within LATAs. If local and long distance are separate but complementary goods like cars and gasoline, why is it such an artificial matter to determine where one ends and the other begins?

The arbitrary nature of the line would not be bothersome if, once it were drawn, it resulted in free competition for homogenous goods. Precisely because it is arbitrary, however, the line will not stand up in an open market but must be guarded by regulation. The MFJ prevents the BOCs from carrying traffic between LATAs; thus, in the name of competition seven of the nation's most likely competitors were removed from the market. Significantly, no such restrictions were necessary in the CPE market once the interface between the network and terminal equipment was standardized. Robust, deregulated competition was possible almost immediately, despite AT&T's overwhelming share of the telephone equipment market.

The divestiture's fictitious boundary will continue to be operative even if Judge Greene allows the BOCs into long distance in the future. The equal access program has created a _de facto_ line of business restriction by arbitrarily designating the LATA as the basis of dividing telecommunications service components. There may be more efficient ways to organize service, but once the system is engineered and regulated to ensure that local exchange carriers cannot integrate their inter- and intra-LATA networking in ways that give them a competitive advantage, there is little to be gained from their entrance into long distance anyway. The attempt to standardize the elements of a nonhomogenous product has a tendency to enforce artificial boundaries between telecommunications markets.

Open Network Architecture is a more radical and in some ways less arbitrary approach to unbundling. Once again, however, the absence of homogeneity poses obstacles to standardization. A network architecture with standardized interfaces for transmission, switching and signalling and completely unbundled service elements can be defined on paper, of course. But the concreteness of real networks prevents uniform application of these criteria. Spatial relationships, the thickness or thinness of traffic, the extent to which traffic from various points can be concentrated, etc., are all decisive influences on the economics of supply. The physical heterogeniety of networks often makes it most efficient to combine functions and service elements. A set of unbundling requirements that makes sense in a New York city central office may be uneconomical in many other locations. No single set of specifications will work for a network that consists of more than one switch. Instead, the process will require a series of judgment calls and continuing regulatory oversight, as did equal access. There are already some disturbing signs that conformance to ONA requirements can result in arbitrary interventions in telecommunications markets.[40]

Conclusion.

Open interconnection may not be the key to unlocking monopoly that it appears to be. As I have shown in this paper, interconnecting rival carriers can act to preempt full-fledged competition in the supply of access by making it unnecessary for carriers to duplicate each other's facilities. Telephone competition in the early 1900s brought the U.S. to the threshhold of universal service precisely because the contestants were not connected with each other. Current policy, in contrast, fosters competition in high-density business routes only, where markets are undersupplied or overpriced due to rate averaging. Increasingly open interconnection policies allow new companies to leave the task of providing universal service to the established network, thereby making it highly unlikely that the benefits of competition will ever reach the bulk of the population.

The root of the problem is our failure to understand the unique economic characteristics of networks. Economic analysis derived from markets composed of identical, homogenous goods have been applied as if by reflex to the telephone system, resulting in an ill-fated attempt to standardize the unstandardizable. Concepts of pricing and competition must be revised to account for networks' inherent lack of homogeneity. Any further attempt to cram the square peg of network economics into the round hole of compatible-component thinking will only result in more regulatory intervention, not less, and a less efficient network.

ENDNOTES

1] See, for example, Comments of Commissioner Eli M. Noam, New York State Public Service Commission, in FCC CC Docket No. 88-2, Phase I:
"For two decades now we have witnessed the erosion of a centralized and uniform monopoly network. ... What is emerging is a system of great institutional, technical and legal complexity which may be best described as a network of networks, serving different regions, user types and software layers. In such an environment, the rules of interconnection of newcomers to the public network becomes perhaps the most important tool of structural regulation. Whoever controls the rules of interconnection controls the network system itself."

2] In the early 1900s, the term "universal service" meant universal _interconnection_ of all subscribers rather than a telephone in every home. The concept was a response to the fragmentation of subscribers caused by the competition between nonconnected telephone companies and the technical limits on long distance transmission. It did not acquire its present implication of 100% subscribership, to be obtained via government-subsidized access, until after World War II.

3] Adams, Walter and James W. Brock, "Integrated Monopoly and Market Power: System Selling, Compatibility Standards and Market Control." _Quarterly Review of Economics and Business_ 22:4 (Winter, 1982), p. 31.

4] Gerald Brock, "Competition, Standards and Self-Regulation in the Computer Industry," in Richard Caves and Marc J. Roberts, eds., _Regulating the Product_ (Cambridge, MA: Ballinger, 1975).

5] James W. Brock, "Market Control in the Amateur Conventional Photography Industry," Ph.D. dissertation, Michigan State University, 1981.

6] Franklin M. Fisher, "Diagnosing Monopoly," _Quarterly Review of Economics and Business_, 19:2 (Summer, 1979); Adams and Brock (1982) _op cit_, p. 35-36.

7] Gerald Brock, _The Telecommunications Industry_ (Cambridge MA: Harvard, 1981) p. 18; Gerald Faulhaber, _Telecommunications in Turmoil_ (Cambridge MA: Ballinger, 1987), p. 121-123.

8] Jeffrey Rohlfs, "A Theory of Interdependent Demand for a Communications Service," _Bell Journal of Economics and Management Science_ 5:1 (Spring 1974) 16-37.

9] See for example, the minutes of the National Telephone Exchange Association Second National Convention, April, 1881, in which the participants discuss the increase in usage and value that occurs with subscriber growth. Practically every public statement of telephone business people from [redacted] 1880 to 1920 included some recognition of "the network externality."

10] Theodore Vail, 1910 *AT&T Annual Report*, p. 39, argues that only monopoly can give universal interconnection.

11] Joseph Farrell and Garth Saloner, "Competition, Compatibility and Standards: the Economics of Horses, Penguins and Lemmings," in H. Landis Gabel, ed., *Product Standardization and Competitive Strategy* (North-Holland, 1987)., p. 7.

12] Joseph Farrell and Garth Saloner, *op cit*, p. 10-11; Paul David, "Clio and the Economics of QWERTY," *American Economic Review*, 75, May 1985, 332-336.

13] See, for example, Comments of Commissioner Noam, *op cit*, p. 5: "The old Bell system achieved predominance by denying its local exchange competitors interconnection to the Bell local networks and to its long lines system, until it was forced to open up." See also Richard Gabel, "The Early Competitive Era in Telephone Communication, 1893-1920," *Law and Contemporary Problems*, 34, 340-359 (1969); Gerald Brock (1981) *op cit*.

14] Data from Bell Labs archives, Warren NJ.

15] A few Bell licensee companies especially hard-hit by competition had surreptitiously initiated sublicensing in 1900, but it did not become a common practice until it was sanctioned as official Company policy in 1902.

16] "...telephone companies established in regions which we do not occupy, if doing a successful business, ...become starting points for attacks upon our system in other places where such opposition is extremely undesirable. In short, if people are willing to venture their own money and do business in a territory we have not occupied, we should regard them and endeavor to have them in fact as allies, and not as competitors." G.W. Leverett, AT&T General Counsel, to Frederick P. Fish, October 17, 1901. Box 1375, Bell Labs archives, Warren NJ.

17] Independent publications such as *Telephony* contain many attacks on and warnings against sublicensing, calling it a "Judas-like scheme." *Telephony*, December 1905, p. 422; June, 1907, p. 368.

18] L.G. Richardson to Theodore Vail, February 27, 1908, Box 1357, Bell Labs archives, Warren, NJ.

19] Ibid.

20] "Gradually, the terms [of the sublicense contract] were relaxed and Bell officials conceded point after point. The duration of the contract was reduced from 99 years to 5, to 3, and finally to one year. The number of [Bell-manufactured] instruments to be rented was reduced from an exclusive total to a meager 100, or 25, finally to one set. The exclusive connection clause was ignored so that the toll circuits of the various competing companies were tacitly allowed to remain in the same sublicensee switchboard." M.D. Atwater, History of the Central Union Company (Bell Labs archives) p. 136-7.

21] U.S. Telephone Company, Annual Report, December 31, 1908. Box 36, Bell Labs archives.

22] L.G. Richardson, President, Central Union Co., to Theodore Vail, July 3, 1908. Box 1357, Bell Labs archives.

23] "Our plan of having all toll lines entering our city on one switchboard has been so pleasant and satisfactory to our patrons that I think that when the [court order] requiring us to remove them becomes known to our patrons, I would not be surprised if some demonstrations on their part would take place expressing their disapproval of being compelled to go back to the old and unsatisfactory way of having more than one toll station in the city." William Shumaker, President, Butler Telephone Co (Independent) to L.N. Whitney, Central Union Co., December 1, 1908. Box 1357, Bell Labs archives.

24] United States Telephone Co. v. Central Union Telephone Co., et. al. 202 Federal Reporter 66-75 (January 10, 1913).

25] A notable exception is Robert Bornholtz and David S. Evans, "The Early History of Competition in the Telephone Industry," in Breaking up Bell (New York: North-Holland, 1983). Bornholtz and Evans explicitly defend the feasibility and value of competition between nonconnected telephone systems.

26] "A telephone system...must cover with its exchanges and connecting lines the whole country. Any development which is comprehensive must cover some territory which is not, and may never become, profitable in itself but must be carried at the expense of the whole." Theodore Vail, 1909 AT&T Annual Report, p. 23. This grand statement comes after 15 years of independent competition. Prior to competition, Bell refused to establish exchanges in small communities; only 10% of the communities with populations under 5,000 had Bell exchanges in 1895.

27] During the years of Bell-independent rivalry, the telephone penetration rate in the U.S. skyrocketed far beyond that of every other country, going from one telephone for every 225 persons to a telephone in every third household and virtually all business establishments.

28] E.g. Gabel (1969), op cit p. 354: "...had there been full interconnection during the early years of competitive rivalry, it may be hazarded that the structure of the telephone industry would have been more equally balanced."

29] The comments of Datran indicated that that company preferred to assume end to end responsibility by allocating 11 Ghz frequencies to a local distribution service. First Report and Order, Specialized Common Carrier decision, 29 FCC 2d 870 (1970) at 955.

30] MCI commented that "local distribution does not present any truly serious problem and...with flexibility of approach, local loop facilities will be available in one form or another." Ibid at 952. MCI filed a related petition to allocate 38.6-40 Ghz radiofrequencies to a Carrier Distribution Service. The petition was supported by Martin Marietta, Resalab, MITEQ, AT&T and Western Union.

31] MCI v. FCC U.S. Court of Appeals, D.C. Circuit. 561 F.2d 365 (1977).

32] Huber Report, p. 3.9, Table IX.5.

33] Open Network Architecture was sanctioned as the future direction of telecommunications policy in the FCC's Third Computer Inquiry, 104 FCC 2d.

34] Personal interview with Michael Cicciano, NYNEX Corp., White Plains, NY, May 17, 1988.

35] CommunicationsWeek, April 25, 1988, p. 32.

36] Farrell and Saloner (1987), op cit, p. 8.

37] Michael Katz and Carl Shapiro, "Network Externalities, Competition and Compatibility," American Economic Review 75:3 (June 1985) 424-440. See also John T. Wenders, The Economics of Telecommunications, Cambridge, MA: Ballinger, 1987, p. 171-190. Wenders compares the market for local exchange and long distance service to the markets for dog food and bread in order to show that an attempt to subsidize the former from the latter could not work in a competitive market.

38] See Wenders (1987) op cit; Leland Johnson, Competition and Cross-subsidization in the Telephone Industry, Santa Monica: RAND Corporation, R-2976-RCNSF, December, 1982.

39] The toll-local switching hierarchy is a more efficient way to organize traffic and network resources, not a boundary between separate products. The whole county could be served by one exchange if we were willing to pay for extremely long access lines. Switching conserves on line facilities and allows trunking economies of scale to be exploited. On the other hand, long distance companies can and sometimes do bypass the local exchange and deliver traffic directly to the doorstep of their customers.

40] A representative of AT&T has written that "To fulfill the Commission's [ONA] requirements, substantial modifications of AT&T's network would have to be made to produce discrete elements with protections that would ensure the technical integrity, reliability and security of AT&T's network. At great expense, AT&T would have to construct new walls and interfaces within the network without knowing whether the demand exists or whether enhanced service providers or end users would be willing or able to pay for these unbundled components." Frank Ianna, "Open Networking in a Competitive Environment," Telecommunications, January 1987 p. 57.

THE POLITICAL ECONOMICS OF STATE RESPONSES TO DIVESTITURE AND FEDERAL DEREGULATION IN TELECOMMUNICATIONS

by Roger G. Noll and Susan R. Smart*

The decade of the 1980s provides an excellent opportunity for studying the political forces that shape the development of economic regulatory policies. It is common knowledge that changes in federal telecommunications policy, and especially the most financially significant divestiture ever accomplished under the antitrust statutes, amount to nothing short of a cataclysmic change in the underlying philosophy of government involvement in the industry. Moreover, as the Federal Communications Commission has moved towards deregulating interstate telecommunications services, the largely cooperative federal-state relationship in telecommunications regulation all but dissolved, while the relative influence of state regulation on the development of the industry increased. The result has been that, for the first time since state regulation was adopted more than a half-century ago, state regulators have had to develop comprehensive policies about pricing and competition in the industry.

These developments provide something of a natural experiment for testing hypotheses about the political and economic forces that shape regulatory policies. Until late 1981, while the FCC was clearly moving to increase the roll of competition in the industry, the ultimate consequence of this policy -- the vertical separation of the local telephone companies from AT&T -- was regarded as a long shot. Indeed, during both the late Carter and early Reagan years, the Antitrust Division of the Department of

Justice was under considerable pressure from other influential members of their Administrations to abandon the quest for wholesale divestiture. Both the FCC and the National Telecommunications and Information Administration, while believing that AT&T had violated the antitrust laws, did not support the scope of divestiture favored by the Department of Justice. Instead, they favored a minor divestiture, such as spinning off one large operating company, combined with injunctive relief and more stringent regulatory rules to protect against anticompetitive actions by AT&T in the future. Then, as the Reagan Administration took the reigns of power, pundits first believed that the case would be settled with a whimper, much as the Eisenhower Administration quickly settled the Western Electric case when it took office. When Assistant Attorney General for Antitrust William Baxter promised to "litigate to the eyeballs," it was widely expected that other members of the Reagan Administration, working with Congress, would vitiate the pending case by enacting legislation which would force the case to be dropped. Indeed, this almost occurred in the summer of 1981, as Assistant Attorney General Baxter was forced to put the case on temporary hold as Congress came within a whisker of passing legislation.[1]

The importance of these developments is that when divestiture was announced in the settlement agreement in December 1981, it was a largely unanticipated event that was simultaneously forced upon all the states. At the end of 1981, the FCC's procompetitive policies had not yet had much of an effect on local telephone service and other activities in the domain of state regulators. Moreover, state and federal authorities had managed to retain a largely cooperative relationship in defining the boundaries of state and federal authority, including the allocation of the industry's revenue requirements between them, in large measure because a vertically integrated AT&T managed to work out many of the inherent conflicts between federal and state authorities within its own corporate structure. With AT&T Long Lines and Bell Operating Companies advocating essentially identical policies, compromise and coordination among regulators was more easily accomplished.

One can make a convincing case that the seeds of disruption in federal-state regulatory relationships were sown long before divestiture, and ought to have been apparent to state regulators before December 1981. Perhaps most regulators even recognized that the world had been permanently and dramatically altered a decade earlier when the FCC was allowing competition in long distance, domestic satellites, and customer equipment. But in 1981, whatever the deeply held views of state regulators, state regulatory policy had remained essentially unchanged. Yet these policies could not remain unchanged in the 1980s, for procompetitive policies were becoming financially significant to the local telephone companies, and through divestiture, the policy integrating power of a vertically integrated AT&T quickly disappeared.

By the time divestiture was implemented in 1984, the basic reality facing state regulators contained the following elements. First, the divestiture agreement had prohibited the Bell Operating Companies from participating in a wide variety of competitive markets, which thereby reduced the effective power of state regulators in regulating these markets even in cases where they retained some authority. Second, in implementing divestiture, the FCC had renounced the policy of burdening services in the federal jurisdiction with an ever growing subsidy of services in the jurisdiction of the states. Third, divestiture had imposed the "equal access" requirement on local telephone companies, which had the effect of requiring them to undertake a massive investment plan to upgrade local switches in order to facilitate competition in long distance. Together these new federal policies undermined the historical pricing policies of the states. Requirement #1 reduced the current and prospective profits of local telephone companies from services that were becoming competitive. Requirement #2 reduced subsidy flows from federal to state services. Requirement #3 imposed new costs because it forced premature retirement of switches that could not cheaply be converted to equal access. Thus, prices for monopoly services regulated by the states had to be increased, and relatively quickly.

The policy question facing state regulators in the 1980s was how to apportion the inevitable rate increases. The twin issues to be decided were the pattern of price increases, and decisions about whether to disallow or disadvantage competition where possible for the purpose of sustaining supercompetitive prices in these services so that prices elsewhere could be held down.

The purpose of this paper is to examine the initial pattern of state responses to divestiture and federal deregulation against the template of contemporary conceptual models of the regulatory process. The next section lays out the fundamentals of the theory of regulatory policy as developed by economists and political scientists during the past two decades, and the status of state regulatory policies circa 1980. Then, the paper develops the implications of this theory for state policy changes in the mid-1980s, and examines the consistency of state policies with these predictions. Because the theory predicts different policy responses among the states, the events of the 1980s give new meaning to the federalist concept of a "laboratory of the states." In this case, divestiture and federal deregulation have created a natural experiment for testing theories of regulatory policy.

THEORETICAL FOUNDATIONS AND INITIAL CONDITIONS

Our analysis is derived from the "new institutionalism" in economic models of policy formation.[2] The purpose of this theory is to provide a conceptual framework for relating the policy preferences of citizens to the policy outcomes of government. The core assumptions of the theory are that all participants in the political process -- voters, candidates for office, and other government officials -- rationally pursue personal goals through the political process; however, they do so in an environment with uncertain and costly information about cause-effect relations, the preferences of other participants in the system, and the exact nature of policy actions taken by other participants. Moreover, owing to differences in economic interests, participants in the political process have divergent objectives, so that democratic political processes are inherently unstable. Hence, the job of elected politicians is not just to enact policies, but to create institutions for making and implementing policies that overcome this inherent instability.

Rational pursuit of personal goals by a politician requires success in winning elections. Even if a politician's ultimate payoff from holding office is determined by the outcome of the policy process, politicians need to be elected in order to participate in making and implementing policy. In principle, the necessity to win elections connects policy to citizen preferences; however, the connection is weak, for reasons having to do with the core assumptions. First, if democratic processes are unstable, a variety of policy outcomes can be supported by election outcomes, and the success of a particular policy in one election can be undone in the next. Second, if a large number of citizens participate in an election, no single citizen is likely to have an effect on the outcome, so citizens have no incentive to determine the policy positions of candidates and vote for the one most favorable to their interests.

For the purposes of this paper, the economic theory of politics provides several important predictions about policy formation. Probably the most often cited result is the prediction that economic interests, organized around commonalities in production and income generation, will be disproportionately influential in politics. The reason is that on the income generating side of economic life, citizens tend to be organized in any case, through firms, unions, and trade associations. Moreover, for nearly all citizens, income production is concentrated in a single industry, perhaps even in a single locality, the financial health of which is by far the most significant factor in determining a citizen's overall economic well-being. Hence, citizens care a great deal about policies affecting their primary source of income, and are organized to deliver support to politicians who work to advance their interests. With respect to regulatory policy in particular, the theory's primary prediction is that the most influential group will be established interests in regulated industries. Among the rest, policy will be biased in favor of the customers or suppliers of the regulated industry that have large financial stakes in its regulation and that are either effectively organized for some other reason (such as that they are already regulated themselves) or that can easily become so.

The reason that organized interests are influential in regulatory processes is political, i.e. as voters, volunteers, and contributors to campaigns they can affect the political futures of elected officials and, through them, state regulators. The implication of this observation is that regulatory policies will be more responsive to the interests of groups located in the state. Hence, one would expect, all else equal, for local telephone companies to fare well relative to national firms in the industry with principal places of business elsewhere.

The mechanism by which organized interests influence politics is to solve two problems for their members: imperfect information and powerlessness. Information about the policy actions of politicians is a public good; hence a large group can act collectively to designate one of their members to become informed and transmit the findings to the members. This substantially reduces the per capita costs of acquiring information. Then, if the organized group can use this information to coordinate political participation among its members, it can overcome the powerlessness of "people like us."

An interesting aspect of interest-group theories of political participation is that, for any given election, only one organized group has to serve the function of getting a voter to the polls. Once there, the voter typically will face a much larger array of offices and candidates than is of interest to the mobilizing organization. After a voter has arrived at the polls, it is relatively costless for the voter to proceed to cast votes in these other contests. Because get-out-the-vote activities of political organizations tend to be oriented to candidates for national offices, an important aspect of understanding contests for state offices is how voters become motivated to cast votes for them.

Again, on their own voters have little incentive to become informed about the candidates for these offices. Hence, their votes (if they vote at all) plausibly can be influenced in two ways. First, based solely on their prior knowledge unrelated to a particular office, they might simply respond to crude signals about the candidate, such as party membership or, if voters are generally satisfied with the overall state of policy to the extent it is known to them, the identity of the incumbent. Of course, because incumbents, by definition, already have won an election, they are likely to be benefitted if these are the only factors influencing their reelection campaigns. In general, incumbents benefit if their constituents remain largely ignorant about the issue positions and other characteristics of candidates in their races.

By implication, the best strategy for a candidate seeking election to an office for the first time is to thwart this strategy by the incumbent. The mechanism for doing so is to bear the costs of informing the voter, much as is done by organized interest groups. The task is to overcome the inertia of voters who have arrived at the polls to vote for other candidates for other offices. Moreover, because each voter will still have a sense of powerlessness regarding the impact of a single vote, the amount of information that a candidate can expect a voter to process is relatively small, for powerlessness leads a voter to devote relatively little effort to becoming informed. Thus, the candidate's strategy is to find a small number of captivating issues that the voter is likely to interpret as symbolizing the overall orientation of the candidate and the opponent. These issues have one of two characteristics: they are of substantial significance in their own rights to the voter's assessment of the performance of the office holder, or the position on a single, relatively unimportant issue signals positions on a wider array of issues that in the aggregate are important. In either case, the choice of issues to be raised by a candidate will be oriented towards aspects of policy that are common to a large number of the voters for a particular office. If not, the information provided by the candidate will be irrelevant clutter to most voters, and hence reduce the likelihood that they will evaluate the candidate positively or pay much attention to additional information from that candidate.

State regulatory policy for utility pricing can become a salient issue to the general electorate in this way, as well as through the efforts of organized interests within the industry. The basic monthly price for telephone service is a ubiquitously shared experience, and it could be interpreted as a signal about a wide variety of related policies, even though the costs of telephone service are a tiny part of the budget of a household. However, recalling that a challenging candidate must focus on only a few issues without losing the attention of voters, changes in utility prices must pass a threshold of significance in order to bump some other issue from the list of items to be emphasized in a campaign. Hence, the optimal strategy for an incumbent, seeking to keep issues away from a challenger, is to keep price changes below the threshold of significance for a large enough number of voters so that they cannot be effectively used as a symbolic issue in a campaign. If the incumbent succeeds, the status quo policy is stable, even if a majority of citizens would prefer some other alternative were they to be completely informed.

The preceding analysis provides a framework for understanding the argument made in a classic paper by Paul Joskow. Joskow's observation is that state regulators pay little or no attention to regulating a utility as long as its prices tend to be falling; however, regulatory scrutiny is greater, and pricing rules more likely to be binding, when costs (and hence prices) are increasing. In terms of the preceding argument, the rationale for this behavior is that prices are unlikely to become a salient political issue as long as they are falling. The reason is that it will be too difficult for a candidate for office to make complicated points about overall profits or the structure of prices when, on balance, most people face lower prices at the time of an election than they did at the last election. Thus, in times of falling prices, interest-group politics will dominate regulatory policy, while in times of rising prices the motives behind changes in pricing policy will be more complex, and reflect greater sensitivity to overall profit rates and the prices paid by ordinary consumers.

It is useful to make a distinction between consumers and businesses regarding utility pricing. Like consumers, most businesses do not regard telephone prices as an important element of overall business costs. In industries in which telecommunications services are used intensively, interest-group arguments pertain: they will seek special price breaks for themselves, not overall reductions in prices for everyone. For other business, the signalling aspect of telephone prices is not as straightforward as it is for consumers. Higher telephone prices can be interpreted as signalling a generally favorable stance towards business. Hence, business may interpret a price hike for them (and the debate about it in a campaign) as indicating that the incumbent is inclined to work for their interests on matters of greater importance to them than monthly telephone rates. As long as the price increases for business customers are not substantially out of line with rate changes for residential customers, business users, then, might be expected to accept them. The implication is that the magnitude of price increases for business might well be greater than for residences, reflecting the lower threshold of price changes that can motivate a political reaction by residential customers.

Finally, one institutional feature of state political systems is also potentially of significance in explaining changes in regulatory policy, which is the geographical basis of representation in state legislatures. In the 1960s, the Supreme Court issued a series of decisions regarding the constitutional requirements for legislative representation. In both the U.S. House of Representatives and state legislatures, legislative districts differed considerably in population, and in general the effect was to overrepresent rural constituencies compared to urban areas. The impact of the Court's decisions was that by the early 1970s, House districts within a state and in both chambers of a state legislature had to be virtually identical in population. Only the U.S. Senate, by virtue of its definition in the Constitution, was exempted from this requirement. The result was a substantial shift in the political influence of rural constituencies, and more so in urbanized states than in the federal government.

The effect of redistricting on telephone pricing was delayed by two factors which operated during the 1970s.[3] The first was the rapid rate of technological advance in the industry, and especially in long distance. The second was the Ozark Plan, an agreement among the FCC, state regulators, and telephone companies to impose an ever-increasing implicit tax on long distance service which would be transferred to local telephone companies to offset an ever-increasing fraction of local service costs. The effect of Ozark was to cause the benefits of cost reductions in long distance to be shared among all services, and between state and federal jurisdictions. During the 1970s, the share of local exchange costs covered by revenues from long distance more than doubled. These revenues enabled regulators to avoid significant price increases for basic local exchange service during the 1970s, despite the unusually high rate of inflation during that decade. Hence, as divestiture and federal deregulation arrived, states were left with an abrupt halt to the growth in revenues from long distance to pay for local network costs and a price structure inherited from the 1960s and before, when the political constituency for regulation was much more rural than in the 1980s.

In expositing the political theory of price regulation, no mention was made of the relationship of prices to costs of service. In fact, prior to the early 1980s, service-specific costs played relatively little role in state rate-making. One purpose of divestiture had been to force more efficient pricing, i.e. to cause prices to be more closely related to costs.[4] But such an effect, according to the theoretical model presented here, would have to be an indirect consequence of the changes in federal policy. The role of costs in the politics of price setting is only that total costs constrain the price-setting role of regulators. Since regulation was permitted by the Supreme Court in the late 19th Century, regulators have been required to permit firms to recover their costs of capital, including a "fair" return on investment. Beyond this, the political factors influencing price setting bear no necessary relationship to costs. Instead, regulated prices will simply reflect the relative importance of the political phenomena described in the theory.[5] If divestiture and deregulation were to drive state regulated prices more towards costs, it would be because the constraints on state regulation, arising from the necessity to cover costs and the limitations on the power of state regulators when the federal component of the industry was more competitive, simply gave the states less flexibility to respond to these political influences.

In practice, the extent to which changes in federal policies did constrain state regulators depended on the specific circumstances of a state. Large, heterogeneous states were not as constrained as were the most rural states, for example, because the latter had fewer services available to continue pricing policies that were advantageous to rural customers. The result is somewhat paradoxical. Given that regulators seek first not to impose substantial price increases anywhere, and that large, urbanized states have more opportunities for cross-subsidy within the price structure, the latter are more likely to continue large rural subsidies, simply for reasons of inertia, even though the political basis for such subsidies disappeared with the redistricting requirements of the 1960s.

POLICY RESPONSES IN THE 1980s

The preceding section suggests several hypotheses about the response of state regulated prices to divestiture and deregulation. First, according to the interest-group argument, local telephone companies ought to have fared relatively well. Second, in seeking to minimize the extent to which changes in federal policy forced changes in state regulated prices, the states would be predicted to strive to maintain the status quo with respect to cross-subsidies. This means erecting barriers to competitive services in the state jurisdiction, and using carrier access charges for long-distance companies within the states as a means to retain revenues from intrastate long distance to offset some local service costs. Third, to the extent price increases for basic access services were necessary, the pattern should be a "spread-the-pain" policy across customer classes, in an effort to keep price increases below the threshold for motivating political response. Fourth, reflecting the more ambivalent views of business towards regulated price increases, changes in prices for business services would be expected to be larger than changes in residential prices. Fifth, reflecting the change in political representation that had occurred since the predivestiture price structure was adopted, prices could be expected to increase more rapidly in rural areas than in large cities. Sixth, any connection between these changes in prices and the service-specific costs of local telephone companies would be coincidental, and driven by the new constraints facing state regulators owing to such factors as federal procompetitive policies and the competitive opportunities created by new technology.

Superficially, the first hypothesis would seem to be contradicted by the initial round of state rate hearings after the announcement of divestiture and several facilitating policy changes by the FCC. Local telephone companies responded by proposing massive rate increases for basic local service, nearly all of which were largely denied by state regulators. Nonetheless, the actual profit performance of the divested Bell Operating Companies was very strong. While accounting profits can be notoriously misleading, these figures are consistent with the appreciation in value of BOC common stocks, and the general view that local service companies have done very well, despite the refusal of states to grant all of the rate increases that they sought.

The second hypothesis is confirmed by the initial reactions of most states to the attempt to introduce competition against local telephone companies. Peter Huber's report to the Department of Justice documents the tendency of most states to prevent competition in intraLATA long distance and in many competitive methods of network access.[6] The exceptions seem to be related to the creation of relatively narrow competitive niches to serve intensive business users of telecommunications services, such as the metropolitan area fiberoptic networks in New York City. Again, this is consistent with the interest-group model of regulatory processes, whereby well-organized groups with intense interests in regulatory policy are provided for. But for most customers, competitive alternatives for local services are generally not permitted, or are not feasible given the nature of regulatory restrictions. Examples are the absence of equal access requirements for facilitating intraLATA long distance competition, the presence of "block or pay" rules or similar provisions to prevent intraLATA long distance competition, and restrictions or prohibitions on shared tenant service, whereby groups of residences or small businesses could form a cooperative local system for purchasing access and other services from either local telephone companies or their competitors.

The third, fourth and fifth hypotheses pertain to the pattern of price changes during the 1980s. To examine whether the patterns and trends in rates conform to these expectations, we have collected single-line service rates for business and residential customers for large operating companies. Single-line business service reveals the price structure for small businesses that generally will not want, or have the opportunity, to acquire competitive services. The data were taken from the 1980s series of Exchange Service Telephone Rates, an annual compilation of telephone prices assembled by the National Association of Regulatory Utility Commissioners. Data were collected for the years 1980 through 1986, except for 1984 when the NARUC compendium was not published. From 1980 through 1983, the NARUC data report rates as of June 30 of the year in question; since then, the rates are those that apply as of December 31 of that year. Data were obtained on all states except Alaska and Hawaii, plus the District of Columbia, and on 56 separate companies.

Each operating company has separate rates for each type of service and for each of several size categories of local exchanges. However, the size categories differ from company to company. To facilitate comparisons among states, rates were collected from each company according to a predetermined set of hypothetical exchange sizes. Ten exchange sizes were used, beginning with the smallest reported exchange and the rate for an exchange with 1,000 customer terminals, and ending with a local exchange of one million terminals. The bottom three size categories represent exchanges in small towns and rural areas.

Many states have no large cities or, if they do, do not offer flat-rate basic service in large metropolitan areas. These states do not quote rates for large local exchanges. Hence, sample sizes are larger for rates in small exchanges than in large ones. In addition, some states report rates for exchange sizes that do not correspond to any actual exchange served by a company within its jurisdiction. These rates were included in the sample.

In a few states, the NARUC data pertain to more than one local operating company. Each company within a state is reported as a separate observation, so that more companies are included in the sample than there are states. Nearly all of the local companies in the sample are part of the Bell System, and since 1985 only Bell Operating Companies are included in the NARUC survey. For 1980-83, the sample also includes General Telephone of Illinois, United Telephone of Ohio, and United Intermountain Telephone of

Tennessee. Unfortunately, data were insufficiently complete to permit inclusion of General Telephone of California.

Table 1 reports the average rates for a single-line residential and business service in each year for the ten size categories of local exchanges. These averages are unweighted by the number of exchanges of each size, and so do not represent the average price actually paid by customers nationwide. Instead, the table shows the trends in state decisions about price structure. The last column shows the change in rates between 1980 and 1986 for each size of locality. The last row shows the difference in price in each year between the smallest and largest local exchanges.

The patterns are consistent with the expectations described above concerning the political forces for rate reform. As the decade begins, both business and residential rates are higher in large exchanges, even though large exchanges have lower costs of service. This is especially true for business service, for which rates were about twice as high in large exchanges as in small ones for the entire period before divestiture (through 1983). In all categories of service prices rose dramatically during the 1980s, a trend that was detectable before divestiture was announced. Given the lags in regulatory processes, the 1982 data probably do not reflect any effects of U.S. v. AT&T; the settlement was announced in December of 1981, but the details were not completed for almost another year. Indeed, even the 1983 prices would reflect only a limited response to the effects of divestiture. Yet prices increased substantially during the early 1980s. More than likely, the first few years reflect the general effects of procompetitive policies at the FCC, combined with other factors such as the inflation of the late 1970s and early 1980s, rapid growth in the demand for new services, and investment programs to upgrade the quality of the network.

Another pattern is that with either the business or the residential categories, the magnitudes of price increases were approximately the same for all sizes of exchanges except the very largest, which were increased by substantially less than the rest. The percentage change in prices was generally higher for smaller communities, and the disparity in prices between small and large communities was narrowed. For both business and residential customers, the difference in prices between the largest and smallest communities fell by about one-third. Thus, on a nationwide basis, the trend has been to reduce the extent to which small towns and rural areas are differentially advantaged in the price structure. This pattern is consistent with the expectation that redistricting can be expected to have reduced the relative influence of rural areas. Nonetheless, substantial benefits to smaller communities remain. Prices are still lower in areas with the highest costs. In fact, the narrowing of price differences by exchange size is probably rather small compared to the magnitude of the subsidy to small communities.

Finally, price increases have been substantially larger for business customers than for residential service. The most interesting aspect of this comparison is that in all but the smallest exchanges, business rates are now above all extant estimates of the average cost of service. These data confirm the hypothesis that, as a political matter, price increases for small business are more palatable than increases in residential prices.

One difficulty in interpreting Table 1 is that relatively few states have exchanges with as many as 1 million terminals; hence, the averages for small communities are not strictly comparable to the entries for large communities. Tables 2, 3, and 4 report the same data as in Table 1, but for three categories of states, according to the size of the largest exchange. The results are somewhat different when the states are so categorized.

Table 2 shows prices for states with only small exchanges. As before, rates have generally risen throughout the 1980s; however, the pattern of increases has not produced much of a narrowing of the price differentials between exchanges of different sizes. This is consistent with the first table, in that most of the narrowing in Table 1 is accounted for by smaller price increases in the very largest communities, none of which are located in the states summarized in Table 2. The primary difference between Tables 1 and 2 is that, for the same range of exchange sizes, the smaller states began the 1980s with larger price differentials between the smallest exchanges and exchanges with 100,000 terminals. This difference disappeared for business service by 1985; however, it has persisted for residential customers. Thus, states in which smaller communities are relatively more important have a price structure that is more favorable to residences in smaller communities.

Table 3 shows the same data for states in which the largest exchange contains either 250,000 or 500,000 terminals. Again, the pattern of rate increases is broadly similar across all categories in the largest exchanges. Almost all of the reduction shown in the "Difference" row is accounted for by smaller price increases in the largest exchange (500,000 terminals). For the rest, relative rates have remained almost identically the same.

Table 4, showing the states with the largest exchanges, also exhibits very little change in the spread in rates between the smallest and largest communities. Price increases are essentially equal across the board, regardless of the size of the exchange. Comparing this table with Table 1 produces an important conclusion: virtually all of the reduction in the differences in rates between small and large communities has arisen because prices in states with the largest cities have increased less rapidly than have prices in the states which contain no large communities. Indeed, rates in small communities were lower in the states with no large cities in 1980-81, but were higher by 1985-86. Thus, for states of all sizes, there is only a slight narrowing of the rate differential by community size; however, federal policy, to the extent that it is the force behind the trends in prices, is causing the differential to narrow by pushing up prices more rapidly in states that do not have large cities.

The pattern of price changes in the 1980s is consistent with the third and fourth hypotheses offered at the outset of this section. For each customer class (residence and small business), states have more or less increased prices across the board, as predicted by the "spread-the-pain" view. And increases have consistently been greater for businesses than for residential customers. In small and medium-sized states prices have increased a little less in exchanges with more than 100,000 terminals, as is consistent with the hypothesis concerning the relative decline of the political importance of rural constituencies; however, as of 1986, this effect was relatively small. Finally, in the most populous states, differences in rates between small and large communities do not appear to have changed, and price increases have been lower than in other states. Apparently regulators in these states are under less pressure to raise rates generally, and have responded by preserving the pattern of cross-subsidies within the rate structure. The somewhat surprising result is that rural customers now pay lower prices in the most urbanized states, where their political influence is presumably not as great.

The last bit of data regarding post-divestiture price decisions by the states is the pattern of carrier access charges for interLATA toll within the states, as shown in Table 5. These are the prices charged by local telephone companies for connecting customers to their long distance telephone company for intrastate long distance calls between LATAs. As shown in the table, these charges are substantially lower in the largest states, but the meaning of these data is difficult to ascertain. Generally speaking, interLATA carriers have interconnection interfaces with local exchange carriers only in larger cities. Hence, the shorter millage distances typically connect customers in larger cities to their long distance carrier, whereas the longer distances are for connections to smaller cities or rural areas. In general, the price structure is consistent with the view that smaller states faced a greater rate shock from divestiture and deregulation, and so imposed a bigger surcharge on long-distance interconnection as part of a general "share-the-pain" strategy. The data also are consistent with a tendency for less populous states to impose a greater price increase for longer distance (and, on average, more rural) service. This reflects a constraint imposed by federal policy, which permits long distance carriers to provide bypass lines to their customers. This option is economically far more attractive when the customer is a relatively short distance from the carrier's point of presence in the LATA. Unfortunately, the data provide only weak evidence for these hypotheses, for the observed price differences might reflect only differences in costs. Larger states presumably have larger traffic volumes for interLATA calls, regardless of the length of local transport, and so may achieve greater economies of scale. Unfortunately, pertinent cost information is virtually nonexistent, so that the conclusions drawn from these prices must be regarded as tentative.

The last hypothesis concerning prices is that only through inadvertance will they reflect costs of service, other than through the requirement that total revenues equal total costs. As with transport costs for carrier access, the actual costs of local service are a matter of considerable uncertainty and controversy. A discussion of this debate is beyond the scope of this paper. Suffice to note that the range estimates is very large. For our purposes, we can simply use estimates of the average cost of local service provided by Bell Operating Companies in various rate hearings during the mid-1980s. Average costs can be compared with prices to determine whether a class of customers is, on balance, subsidized. The estimated average monthly cost from BOCs is in the range of $25 to $30 for most states, with highs over $40 in the least densely populated states. The latter data indicate that in larger urban centers, BOC data would show average costs in the neighborhood of $20. These BOC estimates tend to be at the high end of the range of estimates, so that they can be used as a conservative baseline for identifying customer classes that pay more than average cost for access service.

The comparison between these estimates and the price data in Tables 1 through 4 is interesting, because it indicates that for one class of customers -- small businesses in exchanges larger than 50,000 terminals-- divestiture and deregulation may have driven prices away from costs. In these areas, business customers were paying more or less the cost of serving them in the early 1980s (with prices perhaps above average cost in the largest cities); however, by 1986, business customers in all but the smallest areas were paying prices substantially above any estimate of their average costs. Indeed, in localities with more than 250,000 terminals, small business service costs approximately $10 a month more than these estimates of average costs.

Much less can be said about residential service prices. The data in the tables exclude the FCC's customer access charges. If these are added to the prices in the table, residential prices in communities with over 100,000 lines fall into the range of controversy concerning average costs; however, there is no controversy that for the smallest exchange prices were substantially below costs in 1986. Thus, divestiture can be said to have driven residential prices towards costs, but this effect is small compared to the tendency for small-business prices to be pushed above average costs.

Of course, neither of these trends is likely to have much of an effect on the efficiency of the market for access to the telecommunications network. Both residential and small business demand for service is highly insensitive to price. By contrast, carrier access charges for long distance companies, to the extent that they exceed the traffic-sensitive costs of access, do create inefficiencies. Thus, the primary import of the pattern of single-line price differences and their relationship to costs is what they reveal about the politics of allocating cost responsibilities among classes of customers. It is unambiguously clear from these data that business customers in all but the smallest exchange categories are paying more than the average cost of service, and that residential and business customers in small exchanges continue to receive a very large subsidy.

CONCLUSIONS

The first responses of state regulators to the new policy environment created by divestiture and federal deregulation reveal the complexity of the politics of regulatory policy. The returns of the BOCs and the protections against competition that states have given them provide support for the traditional "capture" theory of regulation, but it is apparent that state regulation of telephones does more than help out regulated firms. The special provisions for large users in larger states, along with supercompetitive profits for BOCs, provide support for the influence of organized interests in shaping regulatory policy. But the data on single-line pricing also reveal a pattern of pricing that is responsive to another form of political influence -- the possibility that poor price performance in a regulated industry could be used by a political entrepreneur as a symbol of the overall policy preferences and performance in office of an incumbent politician. The "share-the-pain" pattern of price increases, and the differential increases between residential and business users, are consistent with the response-threshold characterization of the susceptibility of regulatory policy issues to becoming politically salient. Finally, the somewhat larger price increases for rural customers provide weak confirmation of the view that declining representation of rural interests should cause the structure of prices to shift against them. However, this effect is quantitatively very small. Most likely this reflects the fact that the impact of federal policy changes on state regulation has thus far been too mild to force serious reevaluation of the overall pricing policies practiced by the states. Regulatory officials still adhere to the long-standing policy of using telephone prices to redistribute income to rural communities.

Another aspect of state regulation is that it does not show much of a tendency to move prices towards costs of service. This conclusion is tentative, of course, because so little is known about costs; however, one reason so little is known is that state regulators have not been intensely interested in estimating them. Nonetheless, trends in the rate structure suggest a movement of rural and residential prices towards costs, but a movement of urban business prices away from costs.

The prospects are very good for further exploitation by researchers of state decisions about telephone regulation after divestiture to develop a richer model of the politics of regulation. By collecting more data about pricing, policy institutions, and the economic structure of states, a more sophisticated test of political theories of regulation is clearly feasible. It is apparent that the changes in federal policy of the 1980s provide a rich natural experiment for improving our knowledge of the dynamics of regulation by the states.

FOOTNOTES

* Noll is Professor of Economics, Stanford University, and Sunderland Visiting Professor, School of Law, University of Michigan. Smart is a graduate student in economics, Stanford University. The research reported here was supported in part by a grant from the John and Mary Markle Foundation to the Stanford Center for Economic Policy Research. The authors thank Daniel Kelley for his help in providing data on carrier access charges.

1. For an account of the near termination of the case in late 1980 and early 1981, see Temin and Galambos.

2. For more complete developments of this approach to the study of regulation, see Noll [1988] and Magat, Krupnick and Harrington, Chapter 3.

3. For more details on the development of federal and state regulatory relationships during the 1970s and early 1980s, see Noll [1986].

4. For a complete discussion of the relationship between state regulation and the government's case against AT&T, including its relief proposals, see Noll and Owen.

5. The interests which benefit from regulation have reason to prefer efficient pricing, all else equal, because it maximizes the total economic welfare created by the regulated industry that may then be allocated to them. But all else is rarely equal, for the actions that redistribute the most wealth to these interests are not likely to be consistent with efficient pricing. See Noll [1988].

6. For details about the anticompetitive actions of the states after divestiture, see especially Appendix B of Huber's report.

REFERENCES

Huber, Peter, The Geodesic Network, Washington: U.S. Department of Justice, 1986.

Joskow, Paul L., "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation," Journal of Law and Economics 17 (October 1974), 291-327.

Magat, Wesley A., Alan J. Krupnick, and William Harrington, Rules in the Making, Washington: Resources for the Future, 1986.

Noll, Roger G., "State Regulatory Responses to Competition and Divestiture in the Telecommunications Industry," In Ronald E. Grieson (editor), Antitrust and Regulation, Lexington, Mass.: Lexington Books, 1986.

Noll, Roger G., "Economic Perspectives on the Politics of Regulation," In Richard Schmalensee and Robert Willig (editors), Hanbook of Industrial Organization, New York: North-Holland, 1988.

Noll, Roger G., and Bruce M. Owen, "The Anticompetitive Uses of Regulation: U.S. v. AT&T," In John E. Kwoka, Jr., and Lawrence J. White, The Antitrust Regulation, New York: Scott, Foresman, 1988.

Temin, Peter, and Louis Galambos, The Fall of the Bell System, New York: Cambridge University Press, 1987.

Table 1: Rates for Single-Line Service: Averages for All Companies

YEARS

Size of Locality	Pre-Divestiture		Divestiture Plans		Post-Divestiture		Change
(# of Terminals)	1980	1981	1982	1983	1985	1986	1980-86
BUSINESS:							
Smallest	$14.53	$15.23	$17.03	$19.98	$25.18	$26.04	$11.51
1,000	14.71	15.62	17.44	20.31	25.45	26.29	11.58
5,000	16.33	17.26	19.02	21.80	26.68	27.41	11.08
25,000	19.23	20.23	22.08	25.33	29.69	30.34	11.11
50,000	20.97	22.14	24.39	27.54	32.32	32.78	11.81
100,000	22.93	24.02	26.18	29.64	34.47	34.72	11.79
250,000	25.34	26.99	29.25	32.87	38.04	38.26	11.92
500,000	27.79	29.51	30.78	34.43	39.08	38.86	11.07
750,000	27.91	29.20	33.43	35.23	38.61	38.74	10.83
1,000,000	31.55	34.21	34.06	37.56	38.40	38.59	7.04
Difference	17.02	18.98	17.03	17.58	13.22	12.55	-4.47
# of Companies	50	50	51	50	48	48	
RESIDENTIAL:							
Smallest	$ 6.49	$ 6.69	$ 7.42	$ 8.64	$10.68	$10.92	$ 4.43
1,000	6.60	6.82	7.54	8.76	10.78	11.01	4.41
5,000	7.05	7.25	7.97	9.20	11.15	11.36	4.31
25,000	7.84	8.05	8.82	10.11	11.95	12.13	4.29
50,000	8.26	8.54	9.43	10.73	12.60	12.71	4.45
100,000	8.70	8.96	9.85	11.21	12.94	13.07	4.37
250,000	9.40	9.72	10.56	11.72	13.70	13.81	4.41
500,000	9.86	10.30	10.97	12.24	14.05	14.01	4.15
750,000	9.74	10.02	11.33	12.00	12.89	12.96	3.22
1,000,000	9.56	9.94	10.37	11.44	12.81	12.88	3.32
Difference	3.07	3.25	3.05	2.80	2.13	1.96	-1.11
# of Companies	52	53	54	54	51	51	

Source: National Association of Regulatory Utility Commissioners, Exchange Service Telephone Rates, 1980 through 1983, and Bell Operating Companies Exchange Service Telephone Rates, 1985 and 1986.

Table 2: Rates for Single-Line Service: States With No Large Cities

YEARS

Size of Locality (# of Terminals)	Pre-Divestiture 1980	1981	Divestiture Plans 1982	1983	Post-Divestiture 1985	1986	Change 1980-86
BUSINESS:							
Smallest	$13.87	$13.77	$15.44	$18.54	$25.35	$26.73	$12.86
1,000	13.94	14.64	16.14	19.21	27.76	27.14	13.20
5,000	16.66	17.24	18.66	21.04	27.34	28.36	11.70
25,000	19.70	20.58	22.36	24.79	31.55	32.45	12.75
50,000	22.14	23.33	25.90	27.43	35.35	36.20	14.06
100,000	23.14	23.70	25.58	27.74	34.05	35.24	12.10
Difference	9.27	9.93	10.14	9.20	8.75	8.51	-.76
RESIDENTIAL:							
Smallest	$ 5.99	$ 6.05	$ 6.78	$ 7.92	$10.38	$10.78	$ 4.79
1,000	6.23	6.31	6.96	8.16	10.51	10.90	4.67
5,000	7.02	7.05	7.68	8.85	11.21	11.51	4.49
25,000	7.96	8.08	8.83	9.96	12.34	12.65	4.69
50,000	8.56	8.89	9.85	10.78	13.53	13.81	5.25
100,000	8.78	9.06	9.79	11.09	13.11	13.51	4.73
Difference	2.79	3.01	3.01	3.17	2.73	2.73	-.06
# of States	12	13	13	15	13	13	

Source: Same as Table 1.

Table 3: Rates for Single-Line Service: States With Mid-Size Cities

YEARS

Size of Locality	Pre-Divestiture		Divestiture Plans		Post-Divestiture		Change
(# of Terminals)	1980	1981	1982	1983	1985	1986	1980-86
BUSINESS:							
Smallest	$16.33	$17.34	$19.08	$23.76	$29.63	$30.06	$13.73
1,000	16.69	17.72	19.64	24.06	29.88	30.22	13.53
5,000	18.11	19.25	21.08	25.40	30.82	31.10	12.99
25,000	21.39	22.68	24.38	29.10	33.80	33.89	12.60
50,000	23.05	24.36	26.25	31.36	35.20	34.81	12.76
100,000	25.32	26.71	28.79	33.80	38.21	37.91	12.59
250,000	26.97	28.56	30.40	35.28	39.96	40.20	13.23
500,000	30.74	32.74	32.14	37.23	41.09	40.26	9.52
Difference	14.41	15.40	13.06	13.47	11.46	10.20	-4.21
RESIDENTIAL:							
Smallest	$ 7.24	$ 7.43	$ 7.90	$ 9.64	$12.09	$12.25	$ 5.01
1,000	7.36	7.55	8.03	9.74	12.17	12.30	4.94
5,000	7.76	7.96	8.44	10.09	12.38	12.50	4.74
25,000	8.53	8.74	9.22	10.90	13.14	13.17	4.64
50,000	8.94	9.16	9.77	11.56	13.51	13.39	4.45
100,000	9.55	9.75	10.41	12.24	14.23	14.14	4.59
250,000	10.13	10.35	10.85	12.36	14.83	14.82	4.69
500,000	10.90	11.32	11.03	13.04	15.58	15.10	4.20
Difference	3.66	3.89	3.13	3.40	3.49	2.85	-.81
# of States	18	18	18	17	19	20	

Table 4: Rates for Single-Line Service: States With Largest Cities

YEARS

Size of Locality	Pre-Divestiture		Divestiture Plans		Post-Divestiture		Change
(# of Terminals)	1980	1981	1982	1983	1985	1986	1980-86
BUSINESS:							
Smallest	$13.56	$14.33	$16.43	$17.47	$19.99	$20.43	$ 6.87
1,000	13.67	14.44	16.54	17.58	20.22	20.68	7.01
5,000	15.02	15.74	17.96	19.05	21.57	22.14	7.12
25,000	17.72	18.30	20.62	22.54	24.14	24.73	7.01
50,000	19.30	20.06	22.86	24.72	28.22	29.01	9.71
100,000	21.37	22.44	25.06	27.46	30.93	31.13	9.76
250,000	24.74	26.52	29.37	31.12	37.19	37.43	12.60
500,000	27.16	28.52	31.37	33.10	38.11	38.24	11.08
750,000	27.91	29.20	33.43	35.23	38.61	38.74	10.83
1,000,000	32.09	34.84	34.53	36.18	38.86	39.10	7.01
Difference	19.53	20.51	18.10	18.71	18.88	18.67	-.86
RESIDENTIAL:							
Smallest	$ 6.29	$ 6.54	$ 7.56	$ 8.24	$ 9.22	$ 9.31	$ 3.02
1,000	6.34	6.59	7.62	8.29	9.32	9.42	3.08
5,000	6.67	6.89	7.96	8.64	9.63	9.75	3.08
25,000	7.47	7.65	8.74	9.55	10.35	10.46	2.99
50,000	7.84	8.06	9.25	10.10	11.19	11.32	3.48
100,000	8.25	8.51	9.72	10.60	11.62	11.74	3.49
250,000	9.07	9.44	10.59	11.30	12.56	12.70	3.67
500,000	9.61	10.03	11.26	11.96	13.09	13.23	3.72
750,000	9.74	10.02	11.33	12.00	12.89	12.96	3.22
1,000,000	9.87	10.42	10.82	11.48	12.71	12.79	2.92
Difference	3.58	3.88	3.26	3.24	3.48	3.48	-.10
# of States	19	19	20	19	16	15	

Table 5: Carrier Charges for Premium Switched Access for Intrastate InterLATA Toll, 1987

Local Transport Distance in Miles	Average Total Cost of Switched Access per Minute of Use by Size of Largest Metropolitan Area			
	Small	Medium	Large	Average
under 1	$.065	$.062	$.051	$.059
2	.072	.063	.054	.062
6	.072	.066	.054	.062
11	.076	.067	.057	.065
16	.082	.070	.059	.069
21	.082	.070	.059	.069
31	.091	.079	.068	.078
51	.101	.089	.076	.087
101	.111	.096	.081	.094
Total Number of States*	7	15	20	42

* Eight states and the District of Columbia have a single LATA and, hence, no established carrier access charges for intrastate interLATA toll.

Source: MCI Communications, Inc.

The Uniform Settlements Policy in International Telecommunications:

A Noncooperative Bargaining Approach

Daniel P. O'Brien*

Department of Economics
Northwestern University
October 1988

---

*Support of the Alfred P. Sloan Foundation is gratefully acknowledged. I would like to thank Laura Kodres, Steve Matthews, John Panzar, and Dan Sullivan for comments on an earlier draft and for their many helpful discussions. Of course, any remaining errors are mine.

Abstract

I develop a strategic bargaining model to address the effects of the uniform settlements policy (USP) in international telecommunications. The model predicts that the USP is more likely to increase (decrease) access charges in markets where, under the USP, U.S. firms carry more (less) outbound than inbound traffic. This is due to the model's more general prediction that the USP may allow the foreign monopolist to credibly commit to a take-it or leave-it uniform access charge.

1. Introduction

In April 1979, the Federal communications Commission initiated the first of several actions that are radically changing the international telecommunications market place. Viewing prices as excessive in the voice segment of the market, the FCC considered two broad policy options: 1) open formal rate hearings hoping to determine "appropriate" prices or 2) open entry hoping that competition would hold rates down to cost. Their decision was to open entry to the greatest extent possible.[1]

Unfortunately, the welfare properties of a perfectly competitive closed economy cannot be extended to the international telecommunications services market where the foreign half of the network is monopoly controlled. Most foreign governments are likely to maintain these monopolies well into the future with the unfortunate consequence that they will remain "bottlenecks" in the provision of service. Foreign monopolies interact with U.S. firms in two ways. First, they serve as upstream suppliers of "access to the foreign network", an input used in fixed proportions by U.S. firms to produce "calls to the foreign country". Second, they are also downstream demanders of "access to the U.S. networks". Each foreign monopoly controls not only the number of U.S. carriers allowed to access its network, but also the allocation of U.S. bound traffic across U.S. carriers. These advantages place foreign carriers in a relatively strong bargaining position vis-a-vis their U.S. counterparts in the process that determines the division of international revenues. An important question is whether foreign monopolies can use these advantages to raise the price for access to their networks, reduce the price paid for access to U.S. networks, and ultimately increase the price paid by U.S. consumers for international service.

---

[1] "Order in the Matter of Preliminary Audit," Docket No. 20778, released January 29, 1980, pp. 1-2.

To avoid these potential hazards, the FCC has continued to enforce the *uniform settlements policy* (USP), designed to prevent foreign monopolies from playing U.S. carriers against one another to obtain more favorable agreements. The USP requires that 1) access charges paid by all U.S. carriers to a particular foreign monopoly must be equal, and 2) the access charge paid by U.S. carriers to a particular foreign monopoly must be equal to the charge paid by that foreign monopoly to U.S. carriers. Constraint 1 forbids third degree price discrimination and is subsequently referred to as the "price discrimination constraint." Constraint 2 is referred to as the "50-50 division of tolls."[2] The purpose of this paper is to show that, contrary to the views of the FCC, the USP may result in higher access

---

[2] Readers familiar with the international settlements formula should notice that under the USP what I term the "access charge" is termed "one half the accounting rate" in policy discussions. The accounting rate refers to the basic "unit of account" from which carriers in particular country-pair markets determine the access charge they pay. The "division of tolls" determines what share of the accounting rate each country pays. For example, suppose the accounting rate between the U.S. and France is $2.00, and the division of tolls is 75-25 in favor of France. Then France pays the U.S. $.50 per call-minute for access to the U.S. network, while the U.S. pays French Telecom $1.50 per call-minute for access to the French network. In this terminology the USP requires all U.S. carriers to 1) agree to the same accounting rate and 2) agree to a 50-50 division of tolls. It should be clear that there is no loss of generality in switching from one terminology to the other, and I find the notion of an access charge as a "price" much less confusing.

charges, a higher marginal cost for U.S. firms, and therefore, higher rates paid by U.S. consumers for international service. It turns out that these potentially negative consequences are more likely to arise in the voice and telegraph segments of the market. Thus, the FCC's recent decision to enforce the USP in the voice segment may need to be reassessed.

The key to understanding the welfare effects of the USP is to understand how it affects the degree to which the foreign monopolist controls the access charges paid and received by U.S. carriers. Because of the modelling difficulties posed by multilateral bargaining, the usual approach to this kind of problem has been to assume the agent on the thin side of the market (i.e. the foreign monopolist) has all the bargaining power, regardless of the regime. However, access charges are not unilaterally set by the foreign monopolist — they are the outcome of negotiations. It follows that the effects on access charges of the USP depend on how it alters the relative bargaining abilities of foreign and U.S. carriers.

By introducing bargaining between the foreign monopolist and each U.S. carrier I develop a model that endogenously determines the degree to which the foreign monopolist controls access charges. The model explicitly incorporates the threat each firm has to impose costly delays on firms with which it negotiates as well as the threat foreign monopolist's have to reallocate U.S. bound traffic away from carriers refusing its terms. Equilibrium prices reflect how these threats interact with the USP.

In the regime where the USP is not enforced, the price paid for access to the U.S. network is driven down to U.S. marginal cost. This follows from the observation that, since U.S. carriers are not capacity constrained, the foreign monopsonist plays U.S. carriers against one another a'la Bertrand by threatening to divert traffic to lowest price U.S. carrier. In contrast, the monopolist cannot credibly set a take-it or leave-it price for access to its own network. It attempts to play U.S. carriers against one another as long as two or more have not agreed to an access charge. However, it must bargain bilaterally with the last carrier to reach agreement. U.S. carriers correctly anticipate the benefits of being the last to reach agreement (the monopolist can no longer play one buyer against another) and therefore will not accept the price the monopolist would charge if it could set a take-it or leave-it price. Hence, bargaining equilibrium yields a price intermediate between foreign marginal cost and the price the foreign monopolist would unilaterally set.

The USP changes this story in two ways. First, by constraining agreements to be identical it eliminates bargaining in the subgame after an initial agreement has been reached. Hence, each U.S. firm can no longer credibly reject an offer that another U.S. firm accepts. Second, because of the 50-50 division of tolls, the foreign monopolist may be able to use the threat to reallocate U.S. bound traffic in order to obtain a more favorable uniform access charge. This threat is also available when the USP is not enforced, but it carries no force because U.S. firms compete away the profitability of inbound traffic. Under the 50-50 division, however, U.S. inbound traffic is profitable at all uniform access charges above marginal cost. If equilibrium yields an access charge greater than marginal cost, then this threat will be important.

Employing a non-cooperative bargaining model based on that first introduced by Ariel Rubinstein (1982), I show that, under the USP, there is a subgame perfect equilibrium (which is often unique) in which the foreign monopolist sets a take-it or leave-it uniform access charge. Since this charge affects both its revenues from inbound traffic and its marginal cost in the market for its own outbound traffic, the monopolist may set this charge relatively high or low, depending on the relative importance of its inbound and outbound traffic. For the empirically important case in which the demand for U.S. outbound calls is independent of the foreign price, the USP usually raises U.S. welfare when the volume of foreign traffic bound for the U.S. is greater than the volume flowing in the opposite direction. However, when the net traffic flow is outbound from the U.S., as it is in the voice segment of the international market, foreign carriers prefer higher uniform access charges than U.S. carriers. In this case, the USP may increase U.S. firms' marginal cost, leading to higher U.S. prices, lower U.S. revenues, and lower U.S. welfare.

The remainder of this paper is organized as follows. Section 2 surveys the historical development of the USP, arguing that the FCC has erred by interpretating early Commission decisions as precedent for the policy. Two examples are given that suggest the policy may be failing to meet its objectives in the telegraph market. Section 3 develops a noncooperative bargaining model to address this puzzle. Subgame perfect equilibria are characterized in the constrained and unconstrained regimes.

Section 4 compares equilibrium U.S. welfare in constrained and unconstrained regimes. Section 5 argues that the models' predictions are broadly consistent with recent experience in the telegraph and telex segments of the market. Section 6 offers some concluding comments.

2. The Uniform Settlements Policy: An Historical Perspective

There are three segments of the international telecommunications services market motivating this paper: 1) voice (i.e international long distance calls), 2) telex, and 3) telegraph. Although it was monopolized by AT&T until 1985, the recent entry of MCI and Sprint has raised the question of whether the USP should be enforced in the voice segment. In its recent *Order on Reconsideration*, (FCC 1987), the Commission tentatively ruled that the policy applies, though in a somewhat weaker form than in the telex and telegraph markets.[3] In any case, AT&T still controls most of this market so it may be too early to assess the effects of the USP on access charges by looking at the data. However, the USP originated in the telex and telegraph segments where some degree of rivalry has existed for most of this century. A brief examination of this history provides insight into the effects of the

---

[3] See the section 5 for a discussion of how the rules have been relaxed for the voice market.

USP in telex and telegraph as well as suggesting potential effects in voice.

Legal History

The FCC has been very clear about the intended consequences of the USP. They state: "Our primary responsibility is to U.S. users, not U.S. carriers...Thus, while we prefer to see U.S. carriers rather than foreign administrations maximize their revenues through accounting rate [i.e. access charge] actions, our goal is to facilitate the development of a competitive marketplace characterized by lower rates and greater service/carrier options for users" (FCC 1985, 28419). In pursuit of this goal, they have long feared that unrestrained competition among U.S. firms in telex and telegraph markets might do more harm than good. Frequently, this fear has led them to prevent U.S. firms from signing international agreements that they thought were contrary to the public interest. The first example was in 1936 when they refused to allow Mackay Radio and Telegraph to go into operation between the U.S. and Norway. In the proposed contract, the Norwegian PTT (Postal, Telephone and Telegraph Administration) had agreed to route all new U.S. bound telegraph traffic over Mackay's new circuit. The FCC argued:

> "Inasmuch as the [foreign] telegraph administration controls every word of outgoing radiotelegraph traffic, the competing American radio companies would be dependent upon it for their traffic...Each would be interested in increasing its share of the total traffic. To expect the telegraph administration to play the competing companies against each other is simply to expect that the administration will be headed by good business men, loyal to their national interests. To rely upon companies which are bitter competitors not to make concessions to the administration which controls all outgoing radiotelegraph traffic is to provide an exceedingly tenuous basis upon which to rest public interest" (FCC 1936, 599).

The FCC's basic fear was that the Norwegian PTT would "whipsaw" U.S. firms into paying more for access to the Norwegian network while accepting less for access to their own. This, they argued, would put upward pressure on the price of final service.

Two potential threats that foreign monopolies might use have been cited. As in the Mackay case a PTT might threaten to divert (profitable) U.S. bound traffic from one U.S. carrier to another if the first carrier does not agree to new access charges. Alternatively, a PTT might threaten to terminate its operating agreement with any U.S. carrier refusing to accept less favorable terms. The USP was designed to prevent foreign monopolies from using these threats to gain an advantage harmful to U.S. ratepayers. In the words of the FCC, "This Commission has long maintained a policy of uniformity to preclude 'whipsawing' of U.S. carriers by foreign correspondents. The policy protects the interest of the U.S. and, in particular, the U.S. ratepayer from the adverse effects 'whipsawing' can produce" (FCC 1980, 128).

The puzzling feature of this and many similar statements by the FCC is that to my knowledge, they have never described the mechanism by which the USP benefits U.S. ratepayers. Phrases like "whipsawing" and "play U.S. carriers against one another" have not been made precise. The implicit assumption is that forbidding price discrimination and requiring a 50-50 division of tolls prevents foreign monopolists from credibly threatening (or carrying out threats) to take actions adversely affecting U.S. ratepayers. Yet, neither the threat to reallocate outbound traffic nor the threat to terminate an operating agreement is directly affected by either of these constraints. That is, a foreign monopolist constrained by the USP may find it optimal to threaten to terminate operating agreements or adjust U.S. bound traffic of all U.S. firms not agreeing to its terms as long as at least one U.S. firm does agree.[4] Thus, the USP fails to directly attack either mechanism cited by the FCC as a potential tool for foreign monopolies to gain bargaining advantages over U.S. carriers.

To understand the effects of the USP, one must first understand how it constrains both foreign and U.S. carriers in the bargaining process determining equilibrium access charges. Under the USP, bargaining in a typical country-pair market might proceed roughly as follows. Before initiating service, the foreign monopolist negotiates with U.S. carriers, possibly attempting to "play them against one another," to determine an access charge. This initial charge is not affected by the price discrimination constraint, but it must specify a 50-50 division of tolls. When the initial contract is signed the parties in agreement begin service at the agreed upon charge.

Subsequently, the foreign monopolist may allow other U.S. carriers access. At this point the USP again comes into play. Under current FCC policy, any U.S. carrier reaching agreement to operate at access charges in violation of either the price discrimination constraint or the 50-50 division of tolls must file a request for waiver of the USP with the FCC. After reviewing any objections filed by other carriers, the Commission determines whether a waiver of the USP is in the public interest. Citing its "long standing policy of uniformity," it has become standard practice for the Commission to reject waiver requests in both the telex and telegraph markets.[5] Hence, individual carriers proposing non-uniform access charges are usually required to bring agreements into line with existing agreements before being granted the right to begin (or renew) service.

---

[4] This has also been pointed out by Evan Kwerel (1984).

Notice from this description that there are two phases of the bargaining process during which the USP constrains negotiations. In phase one, before any agreements have been reached, the initial agreement is constrained by the 50-50 division of tolls. In phase two, after an initial agreement has been signed, additional agreements must specify the same price as the initial agreement. The fact that the USP impedes access charge modifications favorable to individual U.S. carriers (and possibly U.S. ratepayers) in phase two is not sufficient reason to argue against the policy. On the other hand, the fact that it also prevents modifications unfavorable to individual carriers in phase two is not sufficient to argue in its favor. Arguments for or against the USP should be based on the effects it has on equilibrium access charges and the implied price of final service as determined via the constraints on both phase one and phase two.

The FCC has not provided arguments along these lines. Rather, it bases support of the USP on historical precedent, often citing the Mackay Radio and Telegraph decision as the first official assertion of its beneficial effects. Unfortunately, it is apparent that current FCC policy makers have not carefully read the Mackay decision. Mackay's proposed contract contained the following paragraph, which was sharply criticized by the commissioners present in 1936:

---

[5] See, for example, FCC (1974) and (1977), where TRT Communications, an international telex and telegraph carrier, was not allowed to implement a lower non-uniform access charge for telex service between the U.S. and the United Kingdom; see also FCC (1985), (1986), and (1987) for other recent examples.

> "Neither party during the continuance of this agreement shall, by modification or renewal of existing agreements or otherwise, enter into an agreement with a third party concerning radiotelegraph traffic between Norway and the United States of America upon terms more favorable than those covered in this agreement or its modification" (FCC 1936, 598).

Notice that, by preventing each carrier from agreeing to any other price with a third party, this provision is formally identical to the USP if it specifies a 50-50 division of tolls (as this agreement did). But the 1936 Commission stated: "Clearly, such a provision could not be said to be in the interest of the American public" (p. 598). Hence, the commissioners addressing the Mackay case in 1936 did not favor the USP. In fact, a careful reading of the case shows that Mackay's proposed access charge was identical to the charge already in existence — the uniformity issue did not even arise:

> "Applicant's witnesses testified that...the division of tolls contemplated by the applicant and the Norwegian Administration will be the same as now effective between RCA Communications, Inc., and the Norwegian Administration on their circuits; to wit, an equal division after the deduction of the out-payments of the Administration and the company, respectively...the testimony was that the out-payments would be the same as those now effective between RCA Communications, Inc., and the Norwegian Administration" (FCC 1936, 595).

It appears, then, that any argument supporting the USP based on the Mackay decision is fallacious.

Rather than contract uniformity, the overriding fear of the 1936 Commission was how future access charges might change throughout the industry to divert revenues to Norway:

> "Although applicant testified that it expects the division to be the same as that on the RCA Communications circuit to Norway, the division can be altered by simple agreement between applicant and the telegraph administration...It is impossible to foresee the ultimate maximum of concessions which a company will make in a desperate effort to get or retain traffic. The Commission should not invite such a situation by granting an application on the facts of the present case, and especially where no offsetting benefits to the public have been shown" (FCC 1936, 599).

Indeed, the proposal was rejected because of Commission fears, fears widely held throughout the great depression, that additional entry would bring about "destructive competition" among U.S. carriers competing for inbound traffic and access to the foreign network. The current Commission's error was to misinterpret the 1936 Commission's concern about destructive competition as calling for uniform settlements. But this interpretation is nowhere to be found in the Mackay case.

This mistaken "precedent" does not end with the 1936 decision. In 1951 Mackay applied for the right to carry telegraph traffic in the U.S./Portugal and U.S./Holland markets (FCC 1951). The current Commission argues that "because the terms of the proposed operating agreement were identical to the terms of the agreements already in effect, our concern was assuaged and the application to place another competitor into the market was granted" (FCC 1985, 28420).

Clearly, there is a problem with this logic. In both 1936 and 1951, Mackay's proposed contract was identical to existing contracts. Entry was not allowed in 1936, but it was allowed in 1951. Therefore, it was not the uniform agreement that assuaged the fears of the 1951 Commission. It was, in fact, their observation that "...a national policy in favor of competition as set forth in the antitrust laws [had] been expressly extended to the field of international communications" (FCC 1951, 734).

To this point, this section has argued that the USP fails on two accounts. First, it does not address the threats cited by the FCC that foreign monopolies might use to whipsaw U.S. carriers. Second, the FCC mistakenly asserts the existence of historical precedent establishing the need for the policy. Hence, an understanding of the policy is not to be found either in current or past Commission statements. Before turning to the model, it remains to examine the data to see if anything can be learned there. I consider two recent examples of opportunistic behavior on the part of foreign monopolies. Not surprisingly, each verifies that the USP does not prevent foreign monopolies from exercising the threat to terminate operating agreements or to reallocate U.S. bound traffic across U.S. carriers. More importantly, each suggests that the policy may be failing to meet its stated objective of protecting both U.S. ratepayers and carriers in the telegraph market.

Case 1: The COMTELCA Telegram

In 1983 a consortium of Central American countries known as COMTELCA sent a telegram to each U.S. telegraph carrier announcing that on a given date they would put new higher access charges into effect in the market for telegraph service between the U.S. and each COMTELCA country.[6] In each U.S./COMTELCA market, U.S. outbound traffic exceeded inbound traffic, implying that the proposed change would favor the COMTELCA countries.[7] The threat used was their assertion that "thereafter, they would deal only with those carriers agreeing to the new charge" (FCC 1985, 28422).

---

[6] COMTELCA (Comision Tecnica Regional de Telecomunicaciones) is composed of the telecommunications administrations of Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua. See (FCC 1985, 28421).

Following the waiver request procedure outlined above, Western Union International was the first U.S. carrier to file a petition with the FCC for waiver of the uniform settlements policy to increase the access charge applicable for service to each COMTELCA country. Similar waiver requests were immediately filed by Western Union Telegraph Co. and FTC Communications. Subsequently, FTC Communications, ITT World Communications, MCI International, RCA Global Communications, TRT Communications, and Western Union International — all the U.S. telegraph carriers operating in these markets — sent the COMTELCA administrations a joint telex informing them that the signatories had agreed to the new access charge effective November 1, 1983. Since no U.S. carrier objected to the new charge, the proposed change was allowed by the FCC.

Prior to the change, U.S. carriers paid \$.1773 per word for access to each COMTELCA network. After the change they paid \$.2365 per word. Hence, the increase in the access charge raised each firms marginal cost by 33 percent. While final service prices rose only a few percentage points in each market, the net revenues of all U.S. carriers combined (i.e. net of payments to the COMTELCA countries) fell by 28 percent from 1982 to 1984.[8] It appears, then, that COMTELCA was able to extract revenue from U.S. firms by using one of the threats the USP was designed to address. Since one of the FCC's goals is to protect U.S. carriers from this kind of whipsawing, it appears that the policy did not work in this case.

---

[7] This point is explained more thoroughly in section 4.

Case 2: The CEPT Telegram

In another instance, on August 19, 1983, RCA Global Communications filed a request with the FCC for waiver of the uniform settlements policy in order to raise the access charge for telegraph traffic between the U.S. and 13 CEPT countries.[9] Again the net traffic flow was outbound from the U.S. implying that the increase would favor the foreign carriers. Other U.S. carriers filed similar waiver requests. Each had received a telegram from CEPT that read:

> "If...your agreement cannot be obtained, we will be forced...to reconsider the agreement reached by us up to the present time and we will take measures for a new breakdown of the traffic and therefore a radical change in the infrastructure" (FCC 1985, 28421).

Notice that both the threat to terminate operating agreements and to reallocate outbound traffic was used by CEPT. The access charge paid by U.S. carriers increased from $.1577 per word to $.2365 per word — an increase of 50 percent — on January 1, 1984.

---

[8] While the average price per word for service to Costa Rica declined (I'm not sure why), the price to the other four countries rose by an average of 5 percent from 1982 to 1984.

[9] CEPT (Conference European des Administrations des Postes et des Telecommunications) is composed of 26 European PTTs.

Table 1 shows how this increase in marginal cost may have been reflected in increased prices for final service. It compares the 1982 average price per word (nominal and real) and number of words outbound to each country with the same variables in 1984. The nominal price of service to each country rose, with an increase of over 20 percent to eight of the thirteen countries. The real price of service to all but two countries rose, with an average increase of 8.6 percent across all thirteen countries.

Of course, the conclusion that increased access charges raised final service prices is only true if changes in other variables were not part of the cause. Since the general trend of cable and satellite circuit utilization charges has been downward, it is probably reasonable to assume that other components of marginal cost did not rise.[10] Hence, unless demand turned much more inelastic or U.S. carriers became much more collusive, it can be concluded that the prices rose because of the increase in access charges.

In each of these examples, foreign monopoly action increased the access charge, and therefore U.S. marginal cost, putting upward pressure on the price of final service. Since the threats used were precisely those the USP was designed to address, each represents a case where the USP apparently failed to meet the FCC's objective of preventing foreign monopolies from whipsawing U.S. carriers in ways harmful to both U.S. carriers and ratepayers. Despite these and other apparent failures, the FCC continues to enforce the USP in the telex

---

[10] See Johnson (1986) for a detailed discussion of the decline in circuit utilization charges.

Table 1

Telegraph Price Increase after CEPT Telegram

	1982		1984			%ΔReal
Country	Price	Words	Price	Words	%ΔPrice	Price
Austria	$.1943	1,295,086	$.2426	866,580	24.9	16.2
Belgium	.2135	1,671,614	.2639	1,113,442	23.6	14.9
Finland	.2111	549,489	.2178	418,760	3.2	-5.5
France	.2209	8,208,943	.2567	5,253,587	16.2	5.5
West Germany	.1800	7,678,938	.2239	6,134,936	24.3	15.7
Ireland	.1506	1,249,949	.1526	835,447	1.3	-7.4
Luxembourg	.1970	219,269	.2168	170.261	10.1	1.3
Netherlands	.1866	2,382,800	.2037	1,456,315	9.2	0.4
Norway	.2215	1,139,593	.2649	777,285	19.6	10.9
Spain	.2209	3,134,151	.2557	2,320,076	15.8	7.1
Sweden	.2047	1,708,767	.2540	1,200,908	24.1	15.4
Switzerland	.1860	3,674,308	.2359	2,830,714	26.8	18.1
Yugoslavia	.1889	869,168	.2415	816,043	27.8	19.2

Source: Statistics of Communications Common Carriers, 1982, 1984.
Notes: %ΔReal Price based on CPI from March of each year.

and telegraph segments and recently decided to enforce it (though less stringently) in the voice segment. It is important, therefore, to develop a better understanding of its effects.

3. A Bargaining Model of International Settlements.

The Model

Consider an international telecommunications market in which a foreign monopolist, M, and two symmetric U.S. carriers, firms 1 and 2, bargain over the price each pays for a continuous flow of the intermediate product, *access* to the foreign network. Access is used by each carrier in fixed proportions to produce a continuous flow of the final product, *calls* to the foreign country. The foreign monopolist produces access at constant marginal cost c, outbound calls via firm i's network at marginal cost $c + a_i^U$ where c is the marginal cost of production due to other competitively sold inputs and a_i^U is the price paid for access to U.S. firm i's network. Similarly, U.S. firm i produces access at constant marginal cost $w \leq c$, and outbound calls at marginal cost $w + a_i^M$. The assumption that U.S. marginal cost is no higher than that of the foreign monopolist reflects the belief that private U.S. firms are at least as efficient as their foreign publicly owned counterparts.

Given a vector of negotiated access charges, $(a_1^U, a_2^U, a_1^M, a_2^M)$, firms are assumed to maximize their instantaneous profit flows at every instant on into perpetuity. This assumption reflects the observation that access charges are usually fixed for a long period of time, while decisions regarding the level of service are made continuously. U.S. firms are Cournot duopolists in the market for calls originating in the U.S.; the foreign carrier continuously produces its monopoly level of service. All firms discount the future continuously at the common rate $\delta = e^{-rz} < 1$, where r is the interest rate, and z is the length of time between successive offers in the bargaining game introduced below. The model is one of complete information — demand, cost functions, the discount factors, and all prices are assumed to be common knowledge.

Turning to the demand side of the market, the utility function of a representative U.S. consumer is $W^U(X) + E^U(y) + z^U$ where X measures the number of originated calls, y the number of calls received, and z^u the consumption of a numeraire good. The utility function of a representative foreign consumer is $W^F(y) + E^F(X) + z^F$ where z^F measures consumption of a numeraire in the foreign country. Both functions are assumed to be concave and continuously differentiable.

In the presence of call-externalities there are a variety of possible assumptions about consumer decision problems in each country. One might assume an *independent decision model* in which the decision to initiate a call from one country is made independently of the decision to initiate a call from the other. Alternatively, the call-externality might be fully internalized. That is, if transactions costs are low, parties in both countries might maximize the joint utility of a call and then negotiate some redistribution of the surplus. It is easy to verify that, under the present utility formulation in which the benefits from inbound and outbound calls are additively seperable, either decision model yields a continuous downward sloping demand function in each country that is independent of the foreign price.[11] Let the inverse demand for calls outbound from the U.S. be $P(X) = W_x^U(X)$; let that for calls outbound from the foreign country be $F(y) = W_y^F(y)$.

Welfare Maximization

The FCC has expressed both the desire to maintain "lower rates" for users and to assure that "U.S. carriers rather than foreign administrations maximize their revenues through accounting rate [i.e. accesss charge] actions." This is captured by assuming that the Commission's goal is to maximize the sum of U.S. consumer and producer surplus, which will be termed U.S. welfare.

The USP affects welfare through its affect on access charges. Assuming no other instruments are available, the benchmark outcome to compare with other equilibria is that in which the Commission chooses access charges to maximize welfare subject to monopoly pricing overseas and Cournot pricing in the U.S. Let $X(a^M)$ be the Cournot equilibrium level of service in the U.S., $y(a^U)$ the monopoly level of service overseas. In the independent call model, total U.S. welfare is

---

[11] The assumption that utility functions are additively seperable can be replaced by the assumption that the distribution of cross partials of the utility functions across individuals in each country is symmetric about zero. See Acton and Vogelsang (1988) for some evidence to this effect from the voice market.

(1) $$US(a^M, a^U) = W^U(X(a^M)) + E^U(y(a^U)) - (a^M + w)X(a^M) + (a^U - c)y(a^U).$$[12]

The surplus maximization problem is

(2) $$\max \{US(a^M, a^U) \mid a^M \geq w,\ a^U \geq c\}.$$

The requirement that the access charges be greater than marginal cost in (2) reflects the assumption that carriers can refuse to accept unprofitable inbound traffic. While one can imagine carriers contemplating agreements to continue service at less than marginal cost, there would be incentives for each carrier to reduce inbound traffic by blocking circuits or quality degradation. That is, such agreements would be inherently unstable.

The first order necessary conditions to (2) (letting subscripts denote derivitives) are

(3) $$[W^U_x - (a^M + w)]X_a - X + \lambda = 0$$

(4) $$E^U_y y_a + y + (a^U - c)y_a + \eta = 0$$

and

(5) $$(a^M - w)\lambda = 0,\ (a^U - c)\eta = 0,\ \lambda \geq 0,\ \eta \geq 0.$$

Let the subscript OPT denote access charges that solve problem (2). Since the first term in (3) is negative, it is clear that $a^M_{OPT} = w$.

---

[12] None of the qualitative results change if the call externality is fully internalized.

Whether a^U_{OPT} is greater than or equal to c depends on the whether the marginal profit to U.S. carriers from increasing a^U above c (y(c)) is greater than or less than the marginal reduction in the call externality $(E^U_y(y(c))y_a(c))$. These two possibilities are illustrated in Figures 1 and 2.

Inbound Traffic

In the unconstrained regime (UR) in which neither the price discrimination constraint nor the 50-50 division of tolls is enforced, negotiations over the access charges for inbound and outbound traffic need not be carried out simultaneously. That is, the foreign monopolist and firm i might negotiate the charge paid for access to the foreign network having already agreed upon that paid for access to firm i's network. Alternatively, some U.S. carriers might reach agreement to carry only one-way traffic, in which case there is only one charge under negotiation for that carrier.[13] Hence, unless carriers can commit to link the decisions to originate and receive traffic, each is essentially involved in two bargaining problems to determine two access charges.

Consider first the market for traffic inbound to the U.S. The foreign monopoly controls the allocation of U.S. bound traffic across U.S. carriers and can shift among them almost instantaneously at almost no cost, provided each U.S. carriers's capacity is not constrained. But capacity is readily available, and it appears that firms can increase it very quickly to handle potentially profitable inbound traffic.[14] Except for the fact that there is only one buyer, this is a classic example of Bertrand competition with U.S. firms competing for foreign inbound traffic. It is well known that, if sellers have constant marginal cost, no capacity constraints, and face downward sloping industry demand, the unique Bertrand equilibrium access charge equals marginal cost.

---

[13] In fact, there have been many cases in which U.S. carriers have agreed to originate, but not terminate, traffic.

It turns out that the Bertrand equilibrium can be realized as a subgame perfect equilibrium to a non-cooperative bargaining game similar to that introduced by Ariel Rubinstein (1982). Imagine a bargaining process in which the monopolist meets sequentially with firms 1 and 2. During each meeting a "proposer" is randomly chosen to offer a price; the other responds "accept" or "reject". If the offer is rejected, the monopolist and that firm enter into (or continue in) a state of disagreement, and the monopolist meets with the other U.S. firm next period under the same conditions. If agreement is reached by the monopolist and firm i, service with firm i begins and the monopolist meets with firm j next period. Now, since U.S. carriers both have constant marginal cost, the monopolist and firm j always agree to under cut i's price, and hence i loses all its inbound traffic. In the limit as the time between bargaining rounds approaches zero, i and j are effectively making simultaneous offers for the monopolist's inbound traffic. It can be shown that the limiting subgame perfect equilibrium to this game converges to the Bertrand equilibrium.

---

[14] Firms in the industry have noted that carriers can add additional capacity in a matter of weeks. See "Implementation and Scope..., Report and Order", p. 4742.

Proposition 1. In the unconstrained regime , the equilibrium price paid for access to the U.S. network is $a_{UR}^{U} - w$.

Outbound Traffic

Negotiations over the prices paid for access to the foreign network are modelled in the same way as those for access to the U.S. network. In each even (resp. odd) period, before any new agreements have been reached, the monopolist meets with firm i (resp. firm j), where the U.S. party to the first meeting is determined randomly. During this meeting a "proposer" is randomly chosen to offer a price; the other responds "accept" or "reject". If the offer is rejected, the monopolist and that firm enter into (or continue in) a state of disagreement, and the monopolist meets with the other U.S. firm next period under the same conditions. If the offer is accepted at time t, the firm that proposed or accepted immediately begins producing the (downstream) monopoly or Cournot output (depending on whether the other carrier is in a state of disagreement), and the foreign monopolist begins bargaining with the remaining firm at time t+1. In this *subgame*, the monopolist and remaining firm alternate offers each period with the monopolist making the first offer. Upon reaching the second agreement each buyer immediately begins producing its Cournot Nash equilibrium output. The game continues in this way until both U.S.

firms have reached agreement.[15] The extensive form of this game is shown in Figure 3.

Following the usual procedure, a subgame perfect equilibrium (SPE) will be derived by proceeding up each branch of the game tree shown in Figure 3, using backward induction. Consider the subgame after the the monopolist and firm j reach agreement to begin service at a_j^M before firm i has reached agreement. At time t, firm j begins producing the (downstream) monopoly output, $x_m(a_j^M)$, allowing the foreign monopolist to earn (flow) profits $U_m(a_j^M) = (a_j^M - c)\ x_m(a_j^M)$. At time t + 1, the monopolist makes an offer to firm i. Since it is already earning profits from firm j, the payoff to the monopolist from reaching agreement with firm i is the incremental profit obtained by moving from monopoly to Cournot equilibrium in the U.S. Let $x_i(a_i^M,a_j^M)$ be the Cournot equilibrium output of firm i when the monopolist and firm i agree on the price a_i^M. Letting U_i be the incremental profits earned by the foreign monopolist by agreeing with firm i, and π_i the incremental (and absolute) profits of firm i,

$$(6)\qquad U_i(a_i^M,a_j^M) = (a_i^M - c)\ x_i(a_i^M,a_j^M) + (a_j^M - c)\ (x_j(a_j^M,a_i^M) - x_m(a_j^M))$$

and

$$(7)\qquad \pi_i(a_i^M,a_j^M) = P(x_i(a_i^M,a_j^M) + x_j(a_j^M,a_i^M))\ x_i(a_i^M,a_j^M)$$
$$- (a_i^M + w)\ x_i(a_i^M,a_j^M).$$

---

[15] One could allow the monopolist or either buyer to choose to leave the game at any time, but these strategies are strictly dominated by choosing to continue bargaining.

The total profit flow earned by the monopolist is

$$(8) \qquad \tilde{U}(a_1^M, a_2^M) = (a_1^M - c)x_1(a_1^M, a_2^M) + (a_2^M - c)x_2(a_2^M, a_1^M)$$

Given a_j^M, the payoff functions U_i and π_i determine a two-firm bargaining problem where the set of possible payoff vectors is

$$(9) \qquad \{ (U_i(a, a_j^M), \pi (a, a_j^M)) \mid U_i(a \ , a_j^M) \geq 0, \ \pi (a_i, a_j^M) \geq 0 \}.$$

The frontier of this set is assumed to be concave.

Equilibrium In the Market for Outbound Traffic

Under the plausible assumption that the opportunity cost of disagreement between bargaining rounds is small relative to the discounted present value of future profits earned after agreement is reached, it is reasonable to restrict attention to the limiting equilibrium as $z \rightarrow 0$ (i.e. $\delta \rightarrow 1$). Ken Binmore (1982) was the first to recognize that, as $\delta \rightarrow 1$, the equilibrium to the two person alternating offer game (i.e. the subgame after one U.S. firm has reached agreement) approaches Nash's (1950) bargaining solution.[16] That is, the unique equilibrium agreement to the subgame after j begins service at the price a_j^M is given by $T_i(a_j^M)$, where

$$(10) \qquad T_i(a_j^M) = \operatorname{argmax} \left\{ U_i(a, a_j^M) \ \pi_i(a, a_j^M) \mid a \geq c \right\}.$$

---

[16] For a formal proof of this result for the present case see O'Brien (1988).

Since the profit frontier is concave, there is a unique solution to problem (10). This observation insures that there is a unique subgame perfect equilibrium outcome in the subgame after only one U.S. firm has reached agreement. The next step is to use this result to derive equilibrium strategies in the subgames before either U.S. firm has reached agreement.

With symmetric buyers, it is natural to investigate stationary equilibria in which, in every period t prior to an initial agreement, the monopolist plans to offer the same price to both U.S. firms, and each U.S. firm plans to offer the same price as the other U.S. firm. A stationary subgame perfect equilibrium (SSPE) of this type can be derived with the aid of Figure 4, which illustrates the functions $T_i(\cdot)$ and the profit contours of the foreign monopolist in access charge space.

The key idea is that when the time between offers is small (δ is close to one), buyer i rejects offers greater than $T_i(a_i^M)$ if it then expects buyer j to agree to a_i^M next period. For, after rejecting such an offer, i expects to wait a very short time before receiving a lower price, $T_i(a_j^M)$, two periods later. Similarly, the foreign monopolist rejects offers from i less than $T_i(a_j^M)$ if it then expects to agree to a_j^M next period. These two conditions, along with the condition that the agent chosen to propose makes an offer maximizing its profits subject to expecting acceptance, imply that the only SSPE offers are given by a_{UR}^M, the price at which T_1 intersects T_2. This conclusion is summarized in Proposition 2. The formal details of the proof can be found in O'Brien (1988).

Proposition 2. In the unconstrained regime, there is a unique SSPE to the game in which firms bargain over a_1^M and a_2^M. In the limit as $\delta \to 1$, the SSPE prices $a_{UR}^M = a_{1,UR}^M = a_{2,UR}^M$ solve

$$(11) \qquad a_{1,UR}^M = T_1(a_{2,UR}^M), \quad a_{2,UR}^M = T_2(a_{1,UR}^M).$$

Define the access charge at which the foreign monopolist *exercises price leadership* in the market for U.S. originated traffic as

$$(12) \qquad a^L = \text{argmax } \{\tilde{U}(a,a) \mid a \geq c\}.$$

Notice from Figure 4 that, since the curve T_i passes below a^L, the monopolist does not exercise price leadership in bargaining equilibrium. The major force preventing it from doing so is each buyer's ability to credibly reject a high price, expecting to agree to a lower price two periods later. This threat is credible because the last buyer to sign a contract has bargaining power. That is, given an agreement between the monopolist and one buyer, adding the second buyer shifts out the derived demand for the monopolist's intermediate product creating an additional surplus over which the monopolist and second buyer bargain. It should not be surprising that the monopolist generally does not receive all the surplus in this phase of the negotiation, and therefore, that forward looking buyers do not accept inordinately high prices early in the negotiations before either buyer has reached agreement.

The fundamental difference between the markets for outbound and inbound traffic in the unconstrained regime is the following. In the former case, bargaining with a second U.S. firm has the potential to increase the monopolist's profits. Hence, U.S. firms have some bargaining power at all prices. In the latter case, adding a second U.S. firm adds nothing to the monopsonist's profits when $a^U = c$, and hence U.S. carriers have no bargaining power at that access charge. This observation suggests an interpretation of what the FCC might mean by foreign monopolists "whipsawing" or "playing U.S. carriers against one another". Whipsawing occurs in the market for inbound traffic, where U.S. carriers are played one against another as they compete for the right to carry the monopsonist's traffic. But it appears that this does not occur to the same degree in in the market for outbound traffic, where it has just been demonstrated that foreign monopolists do not exercise price leadership. If this interpretation is correct then the USP has to potential to increase welfare via its effect on a^U, but it may increase or decrease welfare through its effect on a^M. This intuition is verified in the next section.

Bargaining Under the USP

The USP alters negotiations in two fundamental ways. First, the 50-50 division of tolls forces the monopolist and each U.S. firm to negotiate access charges for both inbound and outbound traffic simultaneously. Second, the price discrimination constraint eliminates bargaining in the subgame after an initial agreement has been reached.

Both of these differences turn out to fundamentally change the equilibrium.

Under the USP, the foreign monopolist is not able to alter a previous agreement with an individual U.S. firm without changing all agreements simultaneously. This can be captured in the bargaining game by introducing a *telegram stage*, prior to time zero, during which the monopolist sends a telegram to each U.S. firm specifying a new uniform access charge and each U.S. firm's share of inbound traffic. This proposal, (a,ρ_1,ρ_2) where ρ_i is i's proposed share of y(a), is to go into effect at time zero. Each U.S. firm simultaneously accepts or rejects this offer. If any one U.S. carrier rejects the offer, previous agreements are terminated and a bargaining procedure similar to that in Figure 4 begins.

The telegram stage of the game can be thought of as a stylization of the CEPT and COMTELCA telegrams discussed in section 2. The role of the Commission at this stage has usually been to reject proposed modifications of the uniform access charge whenever one or more U.S. firms files objections. The Commission has made it clear, however, that it does not intend to become directly involved in the negotiations. Hence, if firms enter the bargaining stage of the game, then the Commission's only role is to enforce the USP.

Suppose the telegram is rejected by firm j but accepted by firm i. Then the bargaining procedure is identical to that in Figure 4 with two exceptions: 1) firm i is the first to meet with the monpolist, and 2) each proposal exchanged between the monopolist and firm i is a vector, (a_i,ρ_i). If the telegram is rejected by both firms 1 and 2, then the first to meet with the monopolist is chosen randomly. If the monopolist and firm i reach agreement at time t, both begin their monopoly levels of service. The foreign monopolist then meets with firm j at time t+1. Under the USP, their only decision is whether to begin service at the price a_i with firm j reciving the fraction $1 - \rho_i$ of U.S. bound traffic. As j begins service, both U.S. firms begin producing their Cournot equilibrium levels of service.

Consider the threats available to the monopolist when it meets with firm i prior to reaching agreement with firm j. In any equilibrium under the USP, the monopolist is indifferent between which U.S. firm carries its U.S. bound traffic. It exercises its greatest possible threat by threatening to divert as much U.S. bound traffic as possible away from firm i if i's agreement cannot be obtained. It should be clear, therefore, that absent any capacity constraints, the first firm to reach agreement receives all traffic terminating in the U.S.

It may seem rather extreme to assume that one firm can receive all U.S. bound traffic. It turns out, however, that all that is necessary for the argument below is that the first firm to reach agreement receive a share $\bar{\rho} > 1/2$. For expositional convenience it is assumed that $\bar{\rho} = 1$.

Returning to the subgame that begins in period zero after the telegram has been rejected, suppose agreement is reached at time t with firm i to begin service at the access charge a^*. The instantaneous profit flow earned by firm i is

$$(13) \qquad \pi_m(a^*) + I_m(a^*)$$

where $I(a^*) = (a^* - w)y(a^*)$ is the profit earned from inbound traffic and $\pi_m(a^*) = P(x_m(a^*))x_m(a^*) - (a^* + w)x_m(a^*)$. The profit earned by the foreign monopolist is

$$(14) \qquad U_m(a^*) = F(y(a^*))y(a^*) - (a^* + c)y(a^*) + (a^* - c)x_m(a^*).$$

In period t + 1 the monopolist and firm j determine whether to begin service at the same access charge. It is easy to see that, assuming total Cournot equilibrium output is increasing in the number of U.S. firms, the monopolist and firm j will always agree to begin service. The profit flows earned by firm i and the monopolist after the second U.S. firm reaches agreement are

$$(15) \qquad \pi(a^*) + I(a^*)$$

and

$$(16) \qquad U(a^*) = F(y(a^*))y(a^*) - (a^* + c)y(a^*) + (a^* - c)X(a^*)$$

where $\pi(a) = (1/2)[P(X(a))X(a) - (a + w)X(a)]$ is the profit earned by firm i from outbound traffic. The functions U, U_m, $\pi + I$, and π_m are assumed to be strictly concave and continuously differentiable.

Suppose the interest rate is r. Then the discounted present value of the profit stream earned by firm i is

(17) $$\frac{(1-\delta)}{r}\pi_m(a^*) + \frac{\delta}{r}\pi(a^*) + \frac{1}{r}I(a^*),$$

that earned by firm j is

(18) $$\frac{\delta}{r}\pi(a^*),$$

and that earned by the foreign monopolist is

(19) $$\frac{(1-\delta)}{r}U_m(a^*) + \frac{\delta}{r}U(a^*).$$

In what follows statements such as "the monopolist earns $(1 - \delta)U_m + \delta U$" mean that it earns $(1 - \delta)U_m/r + \delta U/r$ in present value. This terminology involves no loss of generality under the assumption of agreements are infinitely lived.

In the negotiations between the monopolist and firm i, the range of possible equilibrium access charges is bounded by the maximum and minimum charges that either firm would unilaterally set. The charge the monopolist would unilaterally set is

(20) $$a^{M*} = \text{argmax } \{U(a) \mid a \geq c\}$$

The charge firm i would unilaterally set if it were to receive all U.S. bound traffic is

(21) $$a^{U*} = \text{argmax } \{\pi(a) + I(a) \mid a \geq c\}$$

Equilibrium under the USP

It is useful to begin by investigating stationary subgame perfect equilibria in which the monopolist offers $a^*(\delta)$ and both U.S. firms offer $\hat{a}(\delta)$ in every period. A necessary condition for these prices to be SSPE is that they solve

$$(22) \qquad g(a^*,\hat{a}) = \underset{a \geq c}{\text{argmax}} \ \pi(a) + I(a)$$

$$\text{s.t.} \quad (1-\delta)\, U_m(a) + \delta\, U(a) \geq \frac{\delta}{2}\left[(1-\delta)\,[U_m(a^*) + U_m(\hat{a})] + \delta\,[U(a^*) + U(\hat{a})] \right]$$

$$(23) \qquad h(a^*,\hat{a}) = \underset{a \geq c}{\text{argmax}} \ (1-\delta)\, U_m(a) + \delta\, U(a)$$

$$\text{s.t.} \quad (1-\delta)\, \pi_m(a) + \delta\, \pi(a) + I(a) \geq \frac{\delta^2}{2}\,[\pi(a^*) + \pi(\hat{a})]$$

and

$$(24) \qquad a^* = h(a^*,\hat{a}), \qquad \hat{a} = g(a^*,\hat{a}).$$

Equation (22) says that the firm i chooses $\hat{a}$ to maximize profits subject to giving the monopolist at least as much in present value as it expects to receive in stationary equilibrium if it rejects $\hat{a}$. Equation (23) reflects similar behavior on the part of the monopolist, and (24) insures that both conditions hold simultaneously.

To reflect the belief that bargaining costs are small relative to the discounted value of profits available, attention is focused on the limiting equilibrium as $\delta \to 1$. The following proposition is prooved in the appendix.

Proposition 3. Under the USP there exists a SSPE to the bargaining stage of the game in which $\lim_{\delta \to 1} a^{*}(\delta), \hat{a}(\delta) = a^{M*}$. Moreover, if i) $w < c$, ii) $y(a^{M*}) \geq X(a^{M*})$, or iii) $y(a^{M*}) < X(a^{M*})$ and

$$\phi = \frac{[\pi_m(c) + \pi(c)]\ [-y(c) + X(c)]}{[-X(c)/2 + y(c)]\ U(c)} \notin [0,1],$$

then this subgame perfect equilibrium is unique.

Proposition 3 says that, under many circumstances, the unique subgame perfect equilibrium to the bargaining stage of the game is one in which the foreign monopolist unilaterally sets the uniform access charge. Let us consider the intuition as to why this is so by considering three possibilities for the relative values of a^{M*}, a^{U*}, and c.

Case 1: $a^{M*} > c$, $a^{U*} > c$. Suppose that a^{*} converges to a price different from a^{M*}. For the monopolist to be indifferent between accepting $\hat{a}$ or waiting for a 1/2 chance at a^{*} and $\hat{a}$ tomorrow, it is clear that $\hat{a}$ also converges to the same price as delay costs become small. But then, since $I(a) > 0$ for all $a > c$, the monopolist can offer a price slightly more favorable than a^{*} and still expect firm i to accept. For by rejecting, firm i loses its entire profit from inbound traffic, which is strictly positive, and therefore is greater than the loss due to accepting a slightly less favorable price.

Case 2: $a^{M*} = c$, $a^{U*} > c$. The argument is similar to that in case 1. For all offers greater than c there exists a slightly lower offer such that firm i accepts for fear of losing its inbound traffic.

Case 3: $a^{U*} = c$, $a^{M*} > c$. This case is more difficult. If $w < c$, then all feasible agreements yield positive profits from inbound traffic to the first U.S. firm to reach agreement. Hence, the arguments of cases 1 and 2 apply. If $w = c$, then the following situation may arise. Each U.S. firm unilaterally sets a low uniform access charge (equal to marginal cost) even if it receives all U.S. bound traffic. But if the access charge equals marginal cost, the monopolist's threat to divert traffic away from U.S. carriers refusing its terms carries no force. Hence, there may be a limiting equilibrium in which both U.S. firms offer and accept no more than c. These strategies can be supported in the limiting equilibrium only if $\phi \in [0,1]$. If $\phi < 0$ then $a^{U*} > c$, which is just case 1. If $\phi > 1$, then U.S. profits are high relative to foreign profits which implies that the foreign monopolist has enough leverage to prevent U.S. firms from offering a price as low as c in equilibrium. It turns out that $X(a^{M*}) \geq y(a^{M*})$ usually (though not always) implies that $\phi > 1$. Moreover, as y(c) falls relative to X(c), ϕ is eventually > 1.

To complete the analysis of the bargaining game, the equilibrium strategies in the telegram stage must be specified. It is easy to see that, if the equilibrium to the bargaining stage given in Proposition 3 of is unique, then the monopolist can offer (a^{M*},1/2,1/2) to both U.S. firms expecting that both will agree. For if firm j says no, firm i knows that it can do no better than it does by accepting the offer and receiving all inbound traffic at time zero. But given that i says yes, firm j also says yes since it then retains some inbound traffic.

When c is a SSPE offer, all prices between c and a^{M*} can be supported as subgame perfect equilibia. Hence, for certain values of the parameters, the bargaining model loses its predictive power. It should be pointed out, however, that when U.S. firms have lower marginal costs than the foreign monopolist ($w < c$), the equilibrium is always unique, since U.S. inbound traffic is profitable at all contemplated access charges. This is almost certain to be the case in the Central American telegraph markets where telegrams are delivered by hand rather than by phone as they are in the U.S.[17] Moreover, if y is small relative to X, then equilibrium is likely to be unique.

Note that the arguments in both cases 1 and 2 are independent of the magnitude of $I(a^{M*})$. That is, a seemingly small threat to divert traffic has a large effect on equilibrium when the time between successive offers is small.

---

[17] See FCC(1985), 28421.

4. The Price, Profit, and Welfare Effects of the USP

The main utility of Proposition 3 is that it allows the welfare effects of the USP to be studied by examining a simple constrained optimization problem. Let us proceed under the assumption that the foreign monopolist sets the access charge, reserving the uniqueness issue for later.

The foreign monopolist solves

$$(25) \qquad \max \{F(y(a))\, y(a) - (a + c)\, y(a) + (a - c)\, X(a) \mid a \geq c\}$$

Recognizing that y(a) is the optimal foreign level of service for any a, the Kuhn-Tucker necessary conditions are

$$(26) \qquad -y(a^{M*}) + X(a^{M*}) + (a^{M*} - c)\frac{\partial X(a^{M*})}{\partial a} + \lambda = 0,$$

$$\lambda \geq 0, \quad a^{M*} - c \geq 0, \quad \lambda(a^{M*} - c) = 0.$$

Since $\partial X/\partial a < 0$, the constraint binds ($a^{M*} = c$) whenever $X(a^{M*}) < y(a^{M*})$. Hence, the USP lowers the uniform access charge to foreign marginal cost in markets where the net flow of traffic is inbound to the U.S. Interestingly, this unambiguously improves welfare over the unconstrained regime whenever $w = c$. (See Figure 5). Hence, there may be a normative justification for the USP in markets where, under the USP, the net traffic flow is inbound to the U.S. Notice that the informational requirements for this policy are very weak. If demands are approximately independent and costs are the same across countries, all the FCC needs to do is observe whether the net traffic flow is inbound to the U.S. when the USP is enforced to determine whether the policy is having a beneficial effect. Note further that, since $a^{M*} = c$, equilibrium under the USP is unique in this case.

For $X(a^{M*}) > y(a^{M*})$ the constraint in (26) no longer binds, and $a^{M*} > c$. Whether a^{M*} is higher or lower than a^{M}_{UR} depends on the relative volumes of U.S. outbound and inbound traffic. To see this, introduce the shift parameter $\alpha \epsilon [0,\infty)$ into the derived demand for access to the U.S. network by writing it as $y(a,\alpha)$, where $\partial y/\partial\alpha > 0$, $y \to \infty$ as $\alpha \to \infty$, and $y(a,0) = 0$. As α approaches zero, foreign outbound traffic falls to zero, and $a^{M*}(\alpha)$ approaches $a^{L} = \text{argmax} \{(a - c)X(a) \mid a \geq c\}$. As α grows large, foreign outbound traffic grows relative to U.S. outbound traffic, and a^{M*} approaches c. If $c \geq a^{M}_{UR}$ for all α, then $a^{M*}(\alpha) > a^{M}_{UR}$ for all α. If $c < a^{M}_{UR} < a^{L}$, the continuity of y and U implies that for some α', $a^{M*}(\alpha') = a^{M}_{UR}$. Since $a^{M*}(\alpha)$ is monotonically decreasing in α,[18] it follows that for all $\alpha < \alpha'$, $a^{M*}(\alpha) > a^{M}_{UR}$.

Recall that for y small enough relative to X, $\phi > 1$, and therefore equilibrium under the USP is unique. Since final product price is an increasing function of the uniform access charge in Cournot equilibrium, these conclusions can be summarized as follows.

Proposition 4. Suppose $w = c$. Then the uniform settlements policy reduces the price of U.S. service whenever, under the USP, the volume of traffic originating overseas is greater than the volume originating in the U.S. When $w < c$ the effects of the USP are uncertain. However, if U.S. inbound traffic is low enough relative to its outbound traffic then the USP raises the price of U.S. service.

---

[18] That is, $\partial a^{M*}/\partial\alpha = (\partial y/\partial\alpha)/U'' < 0$ by the sufficient second order condition to ~~(5.11)~~. (25)

Notice that when $X < y$ and $w = c$ the USP also increases the profits of U.S. carriers. This is because the charge for access to the U.S. network equals marginal cost in either regime, while that paid by U.S. carriers in the unconstrained regime (a^{M}_{UR}) is higher than that paid under the USP (c). On the other hand, there are some values of α for which the USP raises U.S. price but also raises the revenues of U.S. carriers. This follows because when $\alpha = \alpha'$, the USP raises the access charge paid U.S. carriers (from c to a^{M}_{UR}) but leaves the charge they pay unchanged.

This is summarized as

Proposition 5. Suppose $w = c$. Then the uniform settlements policy increases the profits of U.S. carriers whenever, under the USP, the volume of traffic originating overseas is greater than the volume originating in the U.S. When $w < c$ the effects of the USP on U.S. profits are uncertain. However, if U.S. inbound traffic is low enough relative to its outbound traffic then the USP reduces the profits of U.S. firms.

It follows from Propositions ~~5~~ 4 and ~~6~~ 5 that, if $y(a^{M*}(\alpha),\alpha)$ is small enough relative to $X(a^{M*}(\alpha))$, the USP unambiguously reduces U.S.

welfare. This can be seen with the aid of Figure ~~6~~5. When y is relatively large $(\alpha = \alpha_0)$, equilibrium access charges under the USP, $(a^{M*}(\alpha_0), a^{M*}(\alpha_0))$, yield higher U.S. welfare than those in the unconstrained regime, $(a^{U}_{UR}(\alpha_0), a^{M}_{UR}(\alpha_0))$. As y becomes less important (α falls to $\alpha_1 \le \alpha'$), the iso-US contours eventually become negatively sloped. Since $a^{M*}(\alpha_1) > a^{M}_{UR}(\alpha_1)$, the negative slope of U.S. welfare contours implies that $US(a^{M*}(\alpha_1), a^{M*}(\alpha_1)) < US(a^{U}_{UR}, a^{M}_{UR})$.

5 Policy Implications

While the conclusions from Propositions 4 and 5 are not as sharp as one would like, they do give the policy maker some basis for judging the efficacy of the USP. In telex markets U.S. firms carry more inbound than outbound traffic in most foreign markets. Hence, Proposition 4 suggests that, if U.S. and foreign carriers share the same marginal cost, the policy reduces the price of final service; Proposition 5 suggests that it raises the revenues of U.S. carriers. Since European costs are probably fairly close to those in the U.S., this model predicts that the USP often raises U.S. welfare in European telex markets.

Interestingly, the decline in telex access charges from 1971 to the present closely follows a similar decline in satellite utilization charges. These charges are a major component of carriers' marginal cost.[19] In 1971 the per half-circuit charge fell from about $2400 per onth to about $1800. It was constant throughout most of the 1970s, but then fell from $1800 to about $800 between 1977 and 1982. Access charges for service between the U.S. and CEPT countries followed similar patterns. They were initially set at $1.50, fell to $1.125 in 1971, and then declined as follows: July 1977, $1.05; July 1978, $.985; January 1981, $.69. Conditional on the 50-50 division of tolls, the revenue effects of these declines all favored the PTTs in the CEPT countries. Moreover, there were cases where certain U.S. carriers were threatened with the loss of U.S. bound traffic if their agreement could not be obtained. While it is not known whether these declines would have occurred in the absence of the USP, they are broadly consistent with the predictions of the bargaining model.

---

[19] Satellite circuits are leased from *Comsat*, the U.S. signatory to the International Satellite Organization, from whom most international satellite circuits are obtained for telecommunications service. See

The voice and telegraph segments of the market are different stories altogether. Propositions 4 and 5 suggest that the USP may reduce U.S. welfare in these markets, since the volume of traffic from the U.S. to the foreign country typically exceeds the volume flowing in the opposite direction. The likelihood that this occurs grows as U.S originated traffic grows relative to that originated overseas.

Traffic disparities were quite large in markets involved in both the COMTELCA and CEPT telegrams. In the former case, U.S. originated traffic was over three times larger than traffic flowing in the opposite direction for all countries except El Salvador. Similarly, the volume of U.S. outbound traffic was double that of inbound traffic in many of the CEPT countries. Unfortunately, the model does not say how large the disparity must be for the policy to reduce welfare.

Johnson (1987) for more details.

In the *Order on Reconsideration* (FCC 1987), the FCC was careful to point out that "uniformity is not an end in itself." They state that "departures from uniformity are permissible if the particular departure does not conflict with [the] objectives [for fair treatment of U.S. carriers, and low rates for U.S. consumers]" (FCC 1987, 1118). The mechanism allowing non-uniformity to occur is simple. The carrier desiring to operate at a non-uniform rate applies to the Commission for a waiver of the policy, and if the Commission takes no action after some amount of time (60 days for telex and telegraph, 10 days for voice) the application is granted.

While there are numerous examples of strong enforcement in telex (usually leading to lower access charges), the early indications are that the weaker stance adopted toward voice is being born out in practice. Indeed, between 1985, the year MCI and Sprint entered the voice market, and February 1987 the FCC received 37 applications for waiver of the USP for voice service to a total of 61 countries. None of those requests were opposed or even commented upon (FCC 1987, 1126).

If one takes the view that the burden of proof rests on policy makers for demonstrating the need for the USP, the per se approach adopted in telex and the rule of reason approach adopted in voice are consistent with the sharp predictions of the telegram game in telex, its ambiguous predictions in voice. This same view, however, would recommend a rule of reason approach in the telegraph market, an approach that does not seem to have been adopted there.

6. Conclusion

This paper argues that, contrary to conventional wisdom, the uniform settlements policy may allow foreign monopolists to credibly set take-it or leave-it uniform access charges. This can have positive or negative welfare consequences, depending on how closely the goals of the foreign monopolist match those of the FCC. In particular, the policy usually reduces access charges, and therefore marginal cost, in markets where there is more traffic terminating in the U.S. than there is overseas. However, there may be negative welfare cosequences if more traffic terminates overseas than in the U.S. Unfortunately, the predictions are not sharp enough to draw unambiguous policy conclusions.

Perhaps the main achievement of this paper is that it explains apparent instances of price-setting behavior on the part of foreign monopolists. Casual arguments at the Commission seem to suggest that, in their view, the USP makes it more difficult for the foreign monopolist to "whipsaw" U.S. carriers. The model in this paper predicts just the opposite in markets where U.S. originated traffic is much greater than that originating overseas. To the extent that the bargaining game captures some of the important strategic features present in the international telecommunications market, this conclusion demonstrates how the noncooperative approach to bargaining can shed light on problems that previously were not well understood.

References

Acton, J.P. and I. Vogelsang (1988), "Telecommunications Demand Over the Atlantic: Evidence from Country-Pair Data," Paper presented at the International Telecommunications Society Conference, Cambridge, MA, June 1988.

Binmore, K.G. (1986a), "Nash Bargaining Theory I-III", in Binmore, K.G. and P. Dasgupta eds. *Essays in Bargaining Theory*. London: Basil Blackwell.

Binmore, K.G., A. Rubinstein, and A. Wolinsky. (1986), "The Nash Bargaining Solution in Economic Modelling," *Rand Journal of Economics* 17:2, 176-188.

Johnson, L.L. (1986), "Excess Capacity in International Telecommunications: Poor Traffic Forecasting or What?," Working Paper, Santa Monica, The RAND Corp.

Johnson, L.L. (1987), *Issues in International Telecommunications: Government Regulation of Comsat*, Santa Monica, The RAND Corp.

Kwerel, E.R. (1984) "Promoting Competition Piecemeal in International Telecommunications," FCC Office of Plans and Policy Working Paper, no. 13.

Nash, J.F. (1950), "The Bargaining Problem," *Econometrica*, 18, 155-162.

O'Brien, D.P. (1988), "The Uniform Settlements Policy in International Telecommunications: A Noncooperative Bargaining Model of Intermediate Product Third Degree Price Discrimination," Unpublished Doctoral Dissertation, Northwestern University.

Ploman, E.W. (1982), *International Law Governing Communications and Information: A Collection of Basic Documents*. Greenwood Press, Westport Connecticut.

Rubinstein, A. (1982), "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50, 97-109.

Shaked, A., and J. Sutton. (1984), "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model," *Econometrica* 52, 1351-1364.

Sutton, J. (1986), "Non-Cooperative Bargaining Theory: An Introduction *Review of Economic Studies* 53, 709-724.

United States Federal Communications Commission (1985), "Implementation and Scope of the Uniform Settlements Policy for Parallel Routes," *Notice of Proposed Rulemaking*, CC Docket No. 85-204 (Released July 3, 1985).

United States Federal communications Commission (1986), "Implementation and Scope of the Uniform Settlements Policy for Parallel International Communications Routes," *Report and Order*, CC Docket No. 85-204, Rm-4796, FCC 86-30, (Released January 30, 1986).

United States Federal Communications Commission (1987), "Implementation and Scope of the International Settlements Policy for Parallel International Communications Routes," *Order on Reconsideration*, CC Docket No. 85-204 (Released February 19, 1987).

United States Federal Communications Commission (1936), "In the Matter of Mackay Radio and Telegraph Company, Inc.," Docket Nos. 3336, 3337, and 3338. (Decided June 3, 1936).

United States Federal Communications Commission (1951), "In the Matter of Mackay Radio and Telegraph Company, Inc." Docket No. 8777, (Decided February 21, 1951).

United States Federal Communications Commission (1977), "Uniform Settlement Rates on Parallel International Communications Routes," *Memorandum Opinion and Notice of Proposed Rulemaking*, FCC 77-369, Docket No. 21265, (Released June 13, 1977).

United States Federal Communications Commission (1980), "Order in the Matter of Preliminary Audit...," Docket No. 20778, (released January 29, 1980).

Appendix

Proof of Proposition 3.

Let $\bar{U}(a) = (1-\delta)U_m(a) + \delta U(a)$, $\bar{\pi}(a) = (1-\delta)\pi_m(a) + \delta\pi(a) + I(a)$. Equations (22) - (24) can then be rewritten as

$$\text{(A1)} \qquad h(a^*,\hat{a}) = \underset{a \geq c}{\text{argmax}}\ \bar{U}(a) \quad \text{s.t.} \quad \bar{\pi}(a) \geq \frac{\delta^2}{2}[\pi(a^*) + \pi(\hat{a})]$$

$$\text{(A2)} \qquad g(a^*,\hat{a}) = \underset{a \geq c}{\text{argmax}}\ \bar{\pi}(a) \quad \text{s.t.} \quad \bar{U}(a) \geq \frac{\delta^2}{2}[\bar{U}(a^*,\hat{a}) + \bar{U}(a^*,\hat{a})]$$

$$\text{(A3)} \qquad a^* = h(a^*,\hat{a}), \quad \hat{a} = g(a^*,\hat{a}).$$

Existence

Since each carrier's profits are strictly concave, there is a compact, convex region over which the foreign monopolist and each U.S. carrier have a conflict of interest. By the theorem of the maximum (eg. Varian 1984, 327), h and g are continuous functions over this region. Hence, Brouwer's fixed point theorem implies that a solution to equations (A3) exists, and therefore a SSPE exists.
I now consider three possible cases that can occur, and demonstrate that there is a SSPE in which the monopolist exercises price leadership in each case.

Case 1: $a^{M*} > c$, $a^{U*} > c$. Suppose a^* converges to $\bar{a} \neq a^{M*}$. For the constraint in (A2) to hold for δ close to one, it is clear that $\hat{a}$ also converges to $\bar{a} \neq a^{M*}$. Then, since $I(\bar{a}) > 0$ for all access charges between a^{M*} and a^{U*}, the monopolist can offer an access charge slightly more favorable to itself and still expect either U.S. carrier to accept. For, by rejecting, that carrier loses its entire profit from inbound traffic, which is greater than the loss due to accepting a slightly less favorable access charge. Hence, $a^{*}(\delta)$, $\hat{a}(\delta) \rightarrow a^{M*}$ in this case.

Case 2: $a^{M*} = c$, $a^{U*} > c$. The argument is similar to that in case 1. For all $\bar{a} > c$, there exists an offer slightly lower than $\bar{a}$ such that firm 1 accepts for fear of losing its inbound traffic.

Case 3: $a^{U*} = c$, $a^{M*} > c$. By arguments similar to those in cases 1 and 2, it is clear that $a^{*}(\delta)$ and $\hat{a}(\delta)$ do not converge to the interior of the interval between a^{M*} and a^{U*}. I construct an equilibrium in this case in which it converges to a^{M*}. Suppose that in every period the monopolist plans to offer

$$\bar{a}(\delta) = \operatorname{argmax} \{(1 - \delta)U_m(a) + \delta U(a) \mid a \geq c\},$$

and U.S. carriers each plan to offer $a'(\delta)$, given by $a'(\delta) = g(\bar{a}(\delta), a'(\delta))$ (It is easy to verify that such an a' exists). These are SSPE offers if and only if $\bar{a}(\delta) = h(\bar{a}(\delta), a'(\delta))$.

Since the monopolist's profits are monotonic over the range relevant for bargaining, the constraint in (A2) (which binds for δ close to one) implies that $a'(\delta) \rightarrow \bar{a}(\delta)$ as $\delta \rightarrow 1$. Therefore, since $I(\bar{a}(\delta)) > 0$ for all δ close enough to one,

$$(1 - \delta)\pi_m(\tilde{a}(\delta)) + \delta\pi(\tilde{a}(\delta)) + I(\tilde{a}(\delta)) > \frac{\delta^2}{2}[\pi(\tilde{a}(\delta)) + \pi(a'(\delta))]$$

for all such δ. That is, $\tilde{a}(\delta) = h(\tilde{a}(\delta),a'(\delta))$ for δ close to one, verifying that $\tilde{a}$ and a' are SSPE. Clearly, $\tilde{a}(\delta) \to a^{M*}$ as $\delta \to 1$. Q.E.D.

Uniqueness

From cases 1 and 2 above, it is clear that $a^*(\delta)$, and $\hat{a}(\delta)$ both converge to either a^{M*} or c. The easier step, which is ommitted, is to show that when $\phi > 1$, they cannot coverge to c.

Suppose, then, that $\phi > 1$. Assume without loss of generality that $a^{M*} > a^{U*}$. Define the sets

$A_f = \{$ a | there is a SPE to the subgame beginning with the foreign monopolist's offer in which it receives $\bar{U}(a)$ $\}$

$A_u = \{$ a | there is a SPE to the subgame beginning with U.S. firm i's offer in which the foreign monopolist receives $\bar{U}(a)$ $\}$.

Let M_i = supremum A_i and m_i = infimum A_i for $i \in \{f,u\}$. The method of proof is to show that, as $\delta \to 1$, the vectors (M_f,M_u) and (m_f,m_u) both converge to the solution to (A3). Since the limiting solution is unique, this implies that the limiting equilibrium is unique.

Lemma 1. Suppose $a^* \in A_f$ and $\hat{a} \in A_u$. Then $a = g(a^*,\hat{a}) \in A_u$.

Proof: Consider the following strategies in the subgame beginning with firm i's offer: firm i offers $a = g(a^*, \hat{a})$, the monopolist agrees to only those offers yielding it at least $\bar{U}(a)$. If agreement is not reached immediately, then all players continue with the strategies supporting the equilibrium $(a^*, \hat{a})$.

Note that since $\bar{U}(a) \geq (\delta/2)[\bar{U}(a^*) + \bar{U}(\hat{a})]$, which is the payoff expected by the monopolist if it rejects a, the monopolist cannot gain by deviating from his strategy. Note further that, since $\bar{U}$ is concave and increasing, either $a = c$, in which case firm i cannot gain by rejecting, or

$$\begin{aligned} \bar{U}(a) &= \frac{\delta}{2}\,[\bar{U}(a^*) + \bar{U}(\hat{a})] \\ &< \frac{1}{2}\,[\bar{U}(a^*) + \bar{U}(\hat{a})] \\ &\leq \bar{U}(\tfrac{1}{2}a^* + \tfrac{1}{2}\hat{a}) \end{aligned}$$

which implies that $a \leq (1/2)[a^* + \hat{a}]$. Since $\bar{\pi}$ is concave and decreasing,

$$\begin{aligned} \bar{\pi}(a) &> \bar{\pi}(\tfrac{1}{2}a^* + \tfrac{1}{2}\hat{a}) \\ &\geq \frac{1}{2}\,\bar{\pi}(a^*) + \frac{1}{2}\,\bar{\pi}(\hat{a}) \\ &> \frac{\delta}{2}\,[\bar{\pi}(a^*) + \bar{\pi}(\hat{a})]. \end{aligned}$$

The term on the RHS of the last inequality is more than the most firm i could receive by offering $a+\epsilon$, inducing the monopolist to reject.

Hence, firm i has no incentive to deviate, and therefore, $a \in A_u$.

Q.E.D.

Lemma 2. $M_u = g(M_f, M_u)$.

Proof: By Lemma 1, for all $a^* \in A_f$, $\hat{a} \in A_u$, $M_u \geq g(a^*, \hat{a})$. In particular, $M_u \geq g(M_f, M_u)$. Suppose this last inequality is strict. Then there is some $a \in A_u$ such that $a > c$ and $a > g(M_f, M_u)$. But this is impossible since firm i could then propose $a - \epsilon$, where ϵ is small, and yield higher profits to the monopolist than the best it could do by rejecting. Hence, the monopolist would accept $a - \epsilon$, contradicting the assumption that $a \in A_u$. Q.E.D.

Lemma 3. $M_f \geq h(M_f, M_u)$.

Proof: Suppose that $a^* \in A_f$ and $\hat{a} \in A_u$. By an argument similar to that used in Lemma 1, it can be shown that $a = h(a^*, \hat{a}) \in A_f$. Hence, since $M_f \geq a = h(a^*, \hat{a})$ for all such a^* and $\hat{a}$, it is true in particular that $M_f \geq h(M_f, M_u)$. Q.E.D.

Lemma 4. 1) In any SPE to the subgame in which the monopolist makes the first offer to firm i, i receives at least $\bar{\pi}(M_f)$. 2) In any SPE to the subgame in which firm i makes the first offer, i receives at least $\bar{\pi}(M_u)$.

Proof: 1) Suppose firm i receives $\bar{\pi}(M_f) - \epsilon$ where ϵ is small. Then this SPE must be one in which agreement does not occur immediately; otherwise the monopolist would receive more than $U(\bar{M}_f)$, contradicting the definition of M_f. Consider the following potential SPE strategies: the monopolist offers $M_f + \lambda$ where λ is such that $\bar{\pi}(M_f + \lambda) > \bar{\pi}(M_f) - \epsilon$ and firm i agrees to only those prices less than or equal to $M_f + \lambda$. If agreement is not reached immediately, all players continue in the original equilibrium supporting $\bar{\pi}(M_f) - \epsilon$ as the payoff to i in the subgame in which the monopolist makes the first offer. Clearly, it is in the monopolist's interest to make this offer, and firm i receives more by accepting than it does in the original equilibrium. Hence, these are SPE strategies in which the monopolist receives more than $\bar{U}(M_f)$, contradicting the definition of M_f.

2) By offering the price M_u, i can assure itself of receiving $\bar{\pi}(M_u)$. Q.E.D.

Lemma 5. $M_f = h(M_f,M_u)$.

Proof: By Lemma 3, $M_f \geq h(M_f,M_u)$. Suppose this inequality is strict. Then there is some $a \in A_f$ such that $a > h(M_f,M_u)$. This implies that firm i receives at most $\bar{\pi}(a) < (\delta^2/2)[\pi(M_f) + \pi(M_u)]$ which, by Lemma 4, is the least i receives by rejecting the price a. Hence, $a \notin A_f$.

Q.E.D.

Lemma's 1 - 5 demonstrates that (M_f,M_u) solves (A3). It remains to demonstrate that (m_f,m_u) also solves (A3) in the limit as $\delta \to 1$.

Lemma 6. $m_u = g(m_f, m_u)$.

Proof: By Lemma 1, for all $a^* \epsilon A_f$, $\hat{a} \epsilon A_u$, $a = g(a^*, \hat{a}) \epsilon A_u$. In particular, $a = g(m_f, m_u) \epsilon A_u$, and therefore, $m_u \leq g(m_f, m_u)$.

Now, suppose $m_u < g(m_f, m_u)$. This is impossible since the monopolist guarantees himself more by rejecting any such price. Hence, $m_u = g(m_f, m_u)$. Q.E.D.

Lemma 7. Let $m_i(\delta)$ and $M_i(\delta)$ be the infimum and supremum of the set $A_i(\delta)$ conditional on δ. There exists δ' such that for all $\delta \geq \delta'$, $m_f(\delta) \geq h(m_f(\delta), m_u(\delta))$.

Proof: Suppose that for all $\delta' < 1$, there exists some $\delta \epsilon [\delta', 1)$ such that $m_f(\delta) < h(m_f(\delta), m_u(\delta))$. Then there exists a sequence $(\delta_t) \rightarrow 1$ such that $\lim_{t\rightarrow\infty} m_f(\delta_t) \leq \lim_{t\rightarrow\infty} h(m_f(\delta_t), m_u(\delta_t))$. Moreover, for all δ_t in the sequence, and for all $\epsilon > 0$, it must be true that firm i rejects the offer $m_f(\delta_t) + \epsilon$ in the equilibrium supporting $m_f(\delta_t)$; otherwise the monopolist would make such an offer.

Consider the conditions necessary for i to reject the offer $m_f(\delta_t) + \epsilon$. If i believes that the monopolist and firm j reach agreement with probability one then i rejects the offer only if

$$\text{(A4)} \qquad \bar{\pi}(m_f(\delta_t) + \epsilon) \leq (\delta^2/2)[\pi(m_f(\delta_t)) + \pi(m_u(\delta_t))]$$

where the term on the RHS of (A4) is the most i expects to receive by rejecting the monopolit's offer. Observe from Lemma 6 and problem

(A1) that as $\delta \rightarrow 1$, m_f and m_u converge to the same price, say m. Taking the limit of (A4) as $\delta \rightarrow 1$ and rewriting the resulting expression yields

$$\text{(A5)} \qquad \pi(m + \epsilon) + I(m + \epsilon) \ \leq \ \pi(m).$$

for all $\epsilon > 0$. Since $I(m + \epsilon) > 0$, this is a contradiction.

Next suppose that i believes that the monopolist and firm j reach agreement only if j is chosen to propose. Then i rejects $m_f(\delta_t) + \epsilon$ only if

$$\text{(A6)} \qquad \bar{\pi}(m_f(\delta_t) + \epsilon) \ \leq \ \frac{\delta^2}{2}\pi(m_u(\delta_t)) \ + \ \frac{\delta^2}{2}[\frac{1}{2}\bar{\pi}(m_f(\delta_t) + \frac{1}{2}\bar{\pi}(m_u(\delta_t))]$$

where the first term on the RHS of the inequality is the most i receives in the event that firm j is chosen to propose and the second term is the most he receives if the monopolist is chosen to propose. Taking the limit as $\delta \rightarrow 1$ yields

$$\text{(A7)} \qquad \pi(m + \epsilon) + I(m + \epsilon) \ \leq \ \pi(m)$$

for all $\epsilon > 0$. But this leads to the same contradiction as above.

Similar arguments to those presented in the last two paragraphs lead to a similar contradiction for the case in which only the monopolist's offer to firm j is accepted in the event that i rejects $m_f(\delta_t) + \epsilon$. Hence it must be the case that for δ close to one, firm i expects the monopolist and firm j to disagree with probability one in the event that i rejects $m_f(\delta_t) + \epsilon$. Since $m_f(\delta_t) > m_u(\delta_t)$ (because $m_u = g(m_f, m_u)$) this implies that

$$\bar{\pi}(m_f(\delta_t)) = \frac{\delta^2}{2}[\bar{\pi}(m_f(\delta_t)) + \bar{\pi}(m_u(\delta_t))].$$

For, if $m_f(\delta_t)$ were higher, it would be refused, if it were lower, the monopolist could raise it and still expect acceptance.

But consider as equilibrium strategies, in the event that i rejects a deviant offer of $m_f(\delta_t) + \epsilon$, the following for the monopolist and firm j: the monopolist offers $m_f(\delta_t) - \gamma$, where γ is small, and j agrees to any price less than or equal to $m_f(\delta_t) - \gamma$. If agreement is not reached, all players continue with the strategies supporting $m_f(\delta_t)$ and $m_u(\delta_t)$ as subgame perfect equilibrium prices. Since $\bar{\pi}(m_f(\delta_t) - \gamma) > \bar{\pi}(m_f(\delta_t)) = (\delta^2/2)[\bar{\pi}(m_f(\delta_t)) + \bar{\pi}(m_u(\delta_t))]$, j accepts this offer. And since $m_f > m_u$, $\bar{U}(m_f(\delta_t) - \gamma) > (\delta/2)[\bar{U}(m_f(\delta_t)) + \bar{U}(m_u(\delta_t))]$ for γ small which implies that the monopolist cannot gain by deviating from his strategy. This implies that $m_f(\delta_t) - \gamma$ is an equilibrium price, contradicting the definition of $m_f(\delta_t)$.

It has been shown that, as δ_t approaches one, each of the possible reasons for i to reject the offer $m_f(\delta_t) + \epsilon$ leads to a contradiction. Therefore, for δ_t close to one, $m_f(\delta_t) < h(m_f(\delta_t), m_u(\delta_t))$ cannot be supported as an equilibrium price. Q.E.D.

It is a straightforward matter to show that Lemmas 6 and 7, along with the fact that (M_f, M_u) solves (A3) imply that $m_f \geq M_f$ and

$m_u \geq M_u$ for δ close to one. Therefore, $m_f = M_f$ and $m_u = M_u$, which implies that the SPE to the telegram game is unique for all such δ.

Figure 1

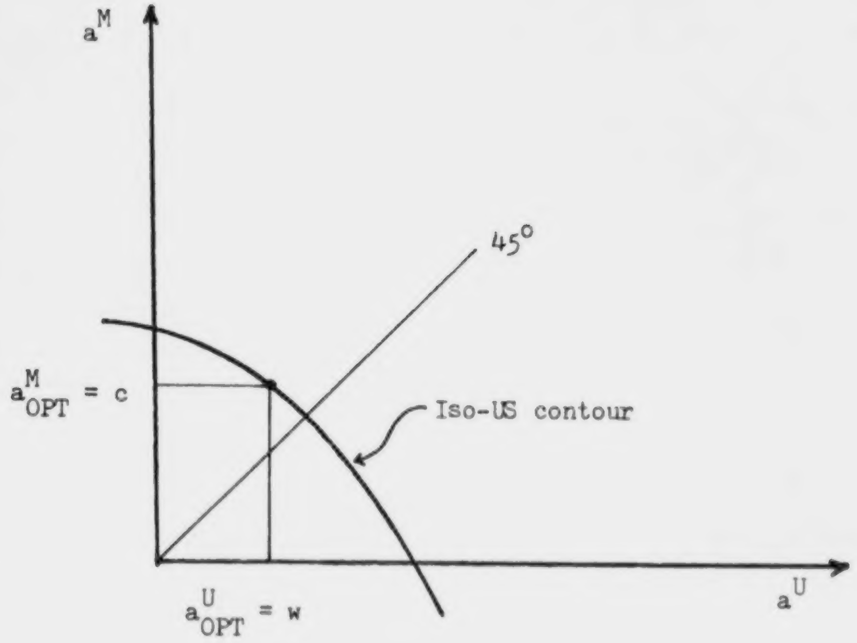

a^M
45°
$a^M_{OPT} = c$
Iso-US contour
$a^U_{OPT} = w$
a^U

Figure 2

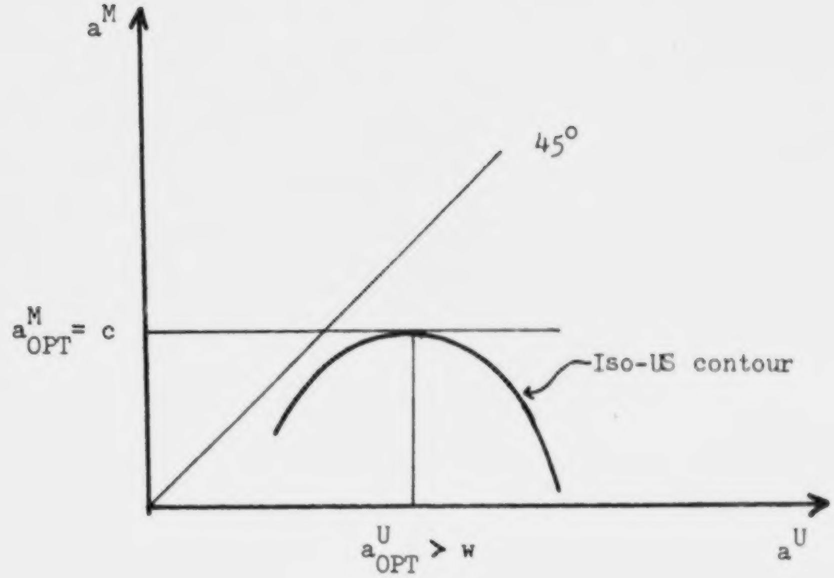
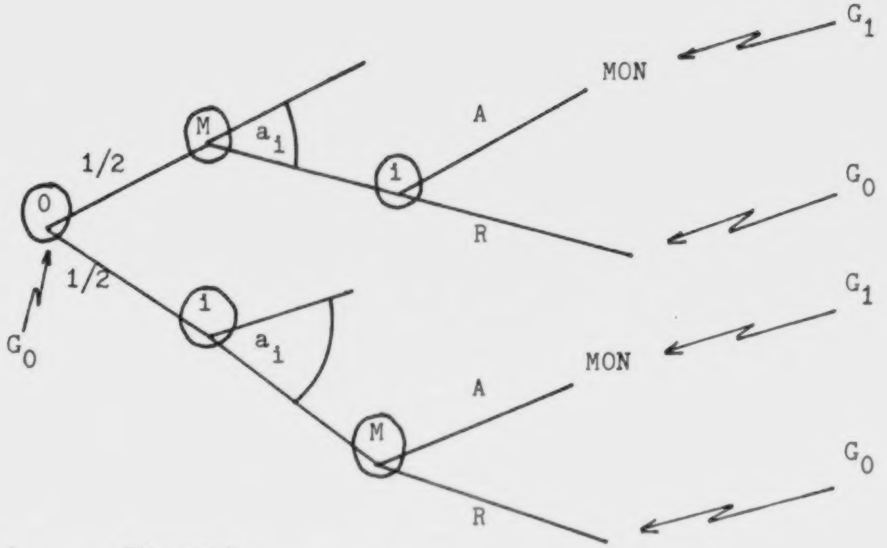

a^M
45°
$a^M_{OPT} = c$
Iso-US contour
$a^U_{OPT} > w$
a^U

Figure 3


G1
MON
A
M
a1
1/2
i
0
G0
R
1/2
i
a1
G1
MON
A
M
G0
R
G0

A = "Accept"
R = "Reject"
0 = "Chance" move
MON = Begin producing monopoly output
CNE = Begin producing Cournot Nash equilibrium output.

The Subgame after buyer i has signed a contract: G_1

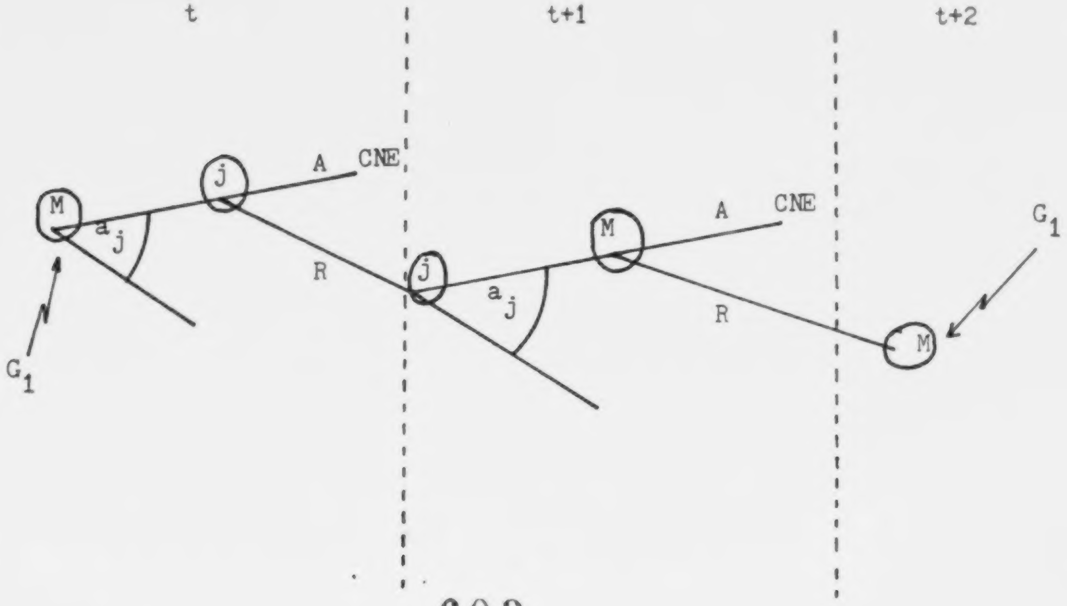

t
t+1
t+2
M
aj
j
A
CNE
R
j
aj
M
A
CNE
R
G1
M
G1

Figure 4

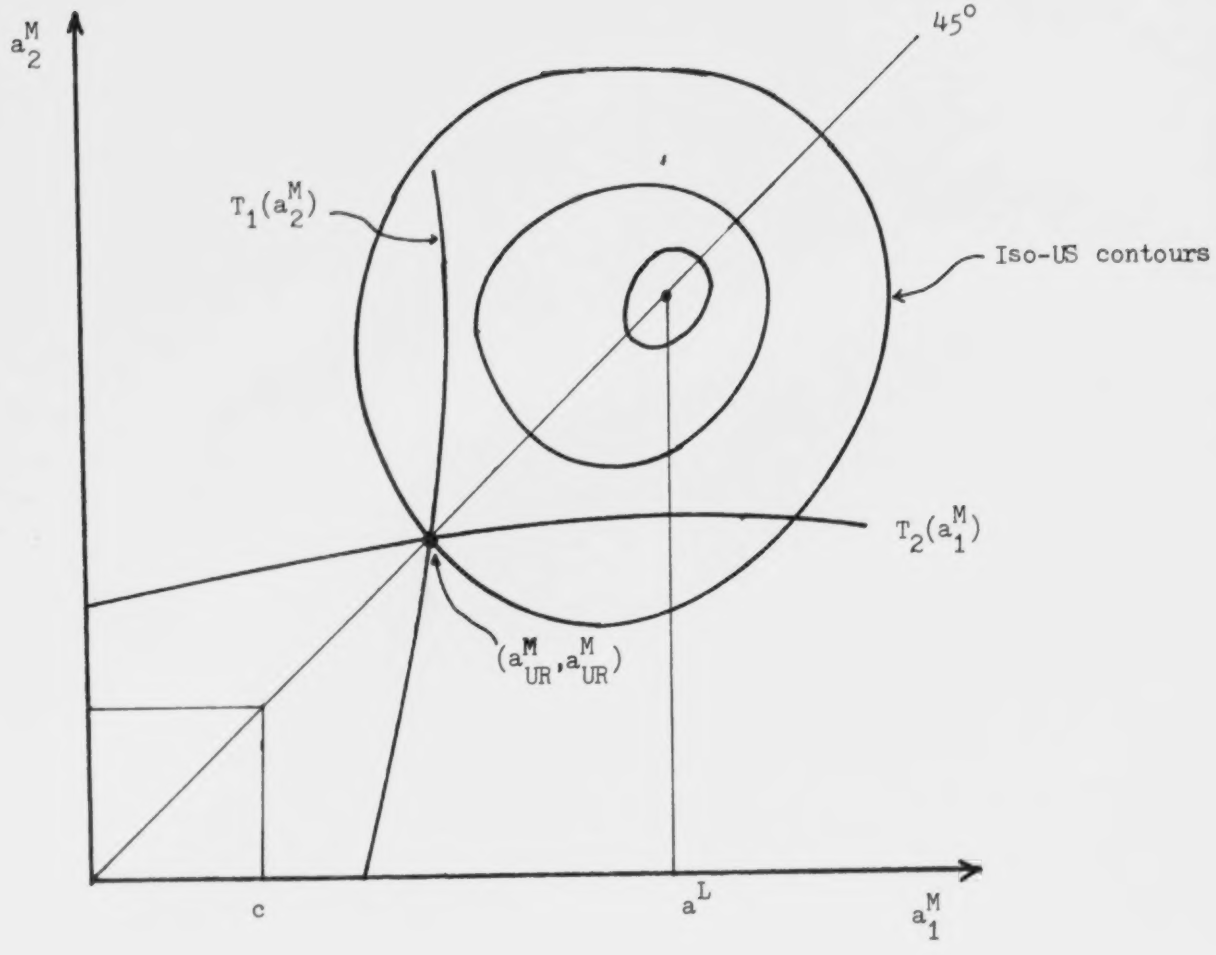
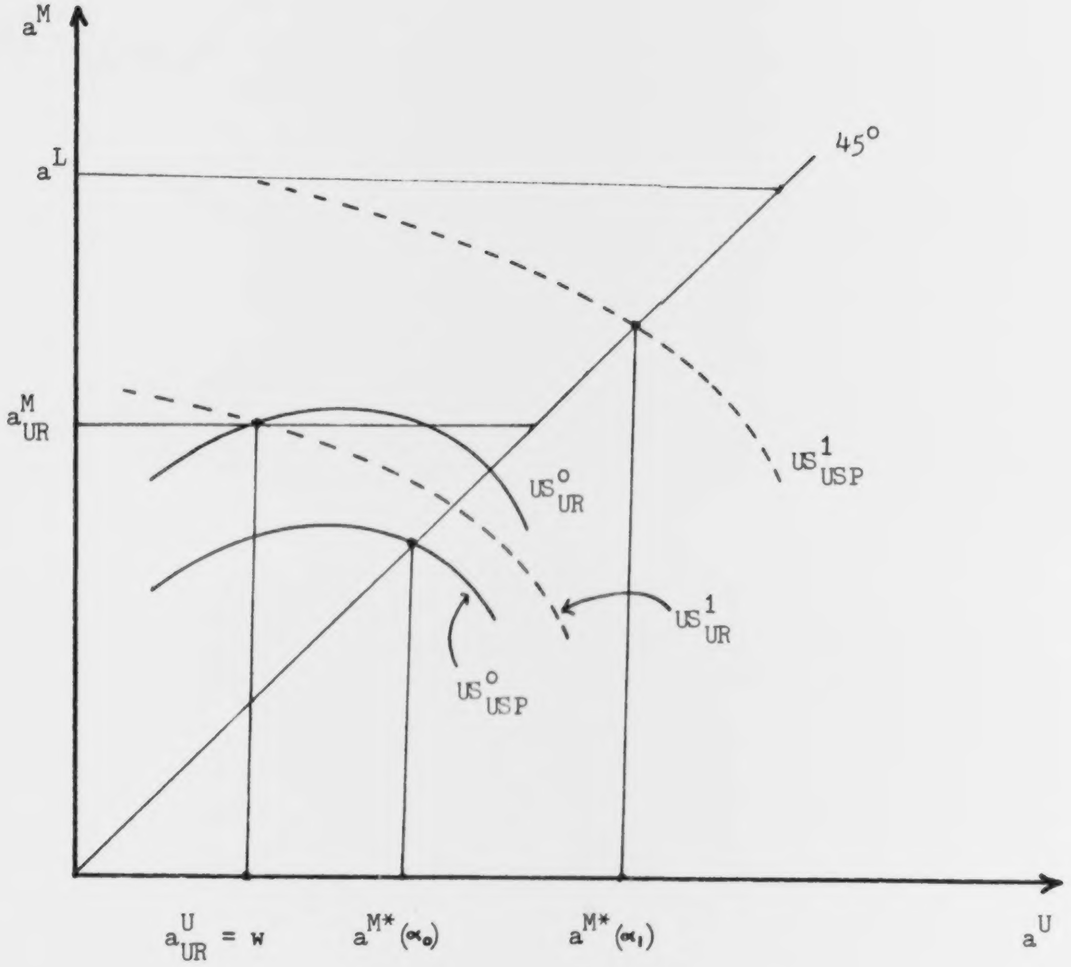

a_2^M
45^o
$T_1(a_2^M)$
Iso-US contours
$T_2(a_1^M)$
(a_{UR}^M, a_{UR}^M)
c
a^L
a_1^M

Figure 5



Notes: US^i_{UR} is the equilibrium Iso-US contour in the unconstrained regime when $\alpha = \alpha_i$.

US^i_{USP} is the equilibrium Iso-US contour under the USP when $\alpha = \alpha_i$.

The Conflict Between Spectrum Efficiency and Economic Efficiency: Diversity Considerations in the Flexible Use of Spectrum Allocated to Radio

Robert P. Rogers and John R. Woodbury

Federal Trade Commission *

October 21, 1988

---

* The views expressed in this paper are those of the authors and do not necessarily represent those of the Commission or any individual Commissioner. That notwithstanding, this paper no doubt contains only the statistically absolute truth. We would like to thank James Langenfeld for his comments on an earlier version of this paper. We are also grateful to David Eaton and Gordon Shuford for research assistance and to Mary Brown for typing assistance.

Introduction

Throughout the entire history of the Federal Communications Commission, one critical component of the agency's decisions regarding the administrative allocation of spectrum to different uses and assignment of spectrum to different users has been the apparent "scarcity" of useable spectrum (Coase (1959)). Because of this perceived scarcity and because the FCC does not charge an explicit price for the use of the spectrum, licensees will not have the correct incentives to conserve the spectrum. Thus the FCC assumes responsibility for ensuring the "efficient" use of the spectrum. However, the agency's definition of spectrum efficiency relies heavily upon spectrum conservation criteria rather than upon maximizing consumer welfare in the use of the spectrum.

Against the background of an economic definition of spectrum efficiency, we consider here one component of this welfare calculus in the context of the programming restrictions placed on the use of commercial FM translators used in providing radio service. In a recent proceeding, the FCC has indicated a reluctance to afford translators greater programming flexibility for spectrum efficiency reasons. However, if a relaxation of the current rules permits more and different choices to listeners and if the additional choices are valued by listeners, then these effects should create a presumption in favor of such flexibility. This case study illustrates the more general conflict between a conservation oriented approach to spectrum efficiency and a consumer oriented approach.

Specifically, we consider whether flexible spectrum use by translators will generate any consumer benefits via greater programming diversity. To infer whether such benefits exist, we ask whether an increase in the total number of radio stations will increase total listenership and whether an increase in the number of radio stations within a programming format will increase the listenership of that format. We also consider whether there is any evidence that consumers are indifferent among the stations offering any particular format.

Based upon a preliminary empirical analysis, we conclude that total listenership does increase with the number of additional radio stations. However this increase results solely from the effect of additional stations on the number of formats offered in the market. The effect of an increase in the number of within-format stations on format listenership is less conclusive. But we cannot reject the hypothesis that there is no listenership impact of increasing the number of within-format stations. Nonetheless, there does appear to be considerable within-format dispersion of listenership shares across stations, suggesting that not all within-format stations are viewed indifferently by listeners. Therefore, restricting the number of stations within a market and within a format will likely artificially limit the benefits consumers receive from broadcast radio.

Clearly, these results fall short of measuring the full net consumer benefits of an additional station (because we lack the metric to place a dollar value on these benefits). But they do suggest that increasing the number of radio stations within a market generates consumer benefits by both increasing the probablility that new formats will be offered and by offering consumers more listening choices within a format. In particular, by permitting translators greater programming flexibility, both of these benefits may be more likely to be attained. Our preliminary results thus indicate that program diversity considerations should carry more weight in the FCC's spectrum use decisions than recently seems to have been the case.

Defining Efficiency in Spectrum Use

Stated in its most general (and therefore empirically empty) way, the economic approach to spectrum efficiency advises that spectrum should be allocated to a particular use until the marginal consumer benefits of that allocation equal its costs. Implementation of this approach requires information on the valuation of the varying amounts of the service by consumers (taking into account the quality of the service provided and any externalities that may be present) and on the costs of supplying the service, including the opportunity cost of the spectrum to be licensed.

In light of the information collection costs associated with the economic approach to spectrum efficiency, many economists have advocated that the FCC be given the authority to auction the spectrum to aspiring licensees (Coase (1959), Kwerel and Felker (1985), Webbink (1981), Mueller) with only limited or no restrictions as to the uses or the transferability of the licenses. In addition to its revenue-generating potential (and a reduction in the expenditure of rent-seeking resources), a competitive auction would result in the assignment of spectrum to the users who anticipate the greatest profits from the use of the spectrum.

For any particular use of the spectrum, auction advocates contend that those profits in general will reflect the value consumers place on the services provided. Similarly, the ability to use the spectrum in the most profitable fashion (i.e., with no use restrictions) insures that over the entire range of spectrum uses, the marginal value to consumers of additional spectrum in any particular use will tend to be equated to the opportunity cost of using that spectrum in any other use. Thus with regard to any particular use and to the entire range of uses, auction advocates argue that consumer welfare would tend to be maximized by a competitive auction approach at any point in time as well as over time as consumer preferences, costs, and technology change.

However, at least with regard to the production of advertiser-supported broadcast services including radio, the auction approach may fail to achieve a first-best welfare maximum (Steiner (1952), Spence and Owen (1977), Borenstein (1988)). First, the profitability of such broadcast stations depends upon the cost of providing programs to listeners and the valuation of those programs by advertisers. Given the array of programming available and the number of stations available to broadcast that programming, the advertiser valuation of the programming depends upon how many listeners are attracted to the programs, not the price consumers would be willing to pay for those programs. As a result, the array of programming offered by advertiser-supported stations will in general differ from that which maximizes listener welfare. The advertiser-supported stations will tend to broadcast programming that attracts many listeners, ignoring those programs valued highly by a relatively small group of consumers. In addition, the programming offered by broadcast stations will tend to be duplicative: Given consumer demands and the station and programming costs, the number of stations offering any particular program format will tend to exceed that which would maximize consumer welfare for that format.[1]

---

[1] This may also be true even if consumers paid directly for the programming, but the extent of duplication will likely be less under a broadcast system financed by listener payments than an advertiser-supported system. See Spence and Owen (1977).

As the number of stations increases, the probability that program formats highly valued by smaller subsets of listeners will be offered increases. But because of the excessive duplication of program formats, the costs of supplying any given array of programs will still be higher than that required for the first-best welfare maximum. And some programs that are highly valued (and whose value exceeds the cost of supplying those programs) may still not be offered to listeners.

Second, there may be other "noneconomic" objectives valued by consumers, ones that are not taken into account by broadcast stations in their profit-maximizing calculus, even if programming were paid for directly by listeners. Most notably, these include goals that foster First Amendment values.

In a world of costless information and a consumer welfare-maximizing regulator, these two difficulties would pose no obstacles to the first-best allocation and assignment of broadcast spectrum. In a world in which both information is costly and the goals of regulators encompasses more than simply maximizing consumer welfare (Peltzman (1976)), the real world choice is between imperfect alternatives to allocate and assign spectrum efficiently.

Even if the market failures associated with broadcast use of the spectrum justify a regulatory allocation and assignment, the approach taken by the FCC differs markedly from the economic approach. While the Commission has never explicitly defined its concept of spectrum efficiency, one recent FCC proceeding regarding FM translators (FCC (1988a)) serves to illustrate at least some of the elements comprising that definition. FM translators are (in effect) low power FM radio stations. Among the components of its regulatory scheme governing translator use, the FCC requires that (commercial) translators receive an off-air FM signal on one frequency and retransmit the signal on a different FM frequency. In addition to the off-air reception restriction, translators are prohibited from virtually all program origination. Together, these restrictions limit the programming choices of the translator licensees to those offered by radio stations geographically located near the translator.

The restrictions on translator uses were imposed to ensure that translators remained a service supplemental to and not a substitute for full power radio stations. The FCC relegated translators to a status secondary to full-power stations because according to the agency full-power stations are more spectrally efficient than translators. In particular, the amount of spectrum required for a group of low-power radio stations to duplicate the coverage area of any full-power station would be greater than that required by the full-power station.

Thus, this concept of spectrum efficiency is strictly an engineering one: In comparing alternative proposals for use of the spectrum, each of which provides broadly similar services (in this case, radio service), the agency will tend to favor licensing the spectrum to the class of users requiring the least amount of spectrum. This particular view of spectrum efficiency has been even more forcefully stated in a recent FCC proceeding (FCC (1988b)) regarding the development of advanced television (ATV) systems: The agency stated that its "preference is for ATV systems that can provide satisfactory service using the least spectrum." Thus, the absolute amount of spectrum used is in and of itself the most important factor in the FCC's "welfare" calculation.[2]

Administrative implementation of the economic approach to spectrum efficiency obviously requires considerably more information than the engineering approach. If the allocation of more spectrum to a particular use results in more or a higher quality service and if the value of that additional service or quality to consumers exceeds the opportunity costs of the additional spectrum and of other inputs, the additional spectrum should be so allocated. The engineering approach is so simplistic that it is unlikely that the correspondence between the "bottom line" of the two approaches is high.

Viewed from the perspective of the economic approach to spectrum efficiency, the issue of permitting translators greater flexibility in programming turns in part on the value consumers would place on the additional programming offered by translators in comparison to the costs of providing that programming. In the empirical work described below, we assess some components of the diversity-related consumer benefits that may flow from permitting the entry of additional radio stations, a situation we consider analogous to permitting greater flexibility in spectrum use by translators. While it may not be possible to completely fill in the blanks in the welfare calculation (and in particular to place a dollar value on those benefits and compare that dollar valuation to the costs of providing the new service), a finding that consumers place a positive value on additional stations would at least require more careful assessment of the consumer benefits that may flow from providing translators enhanced programming flexibility.

---

[2] Of course, if for any given amount of identical services produced, one method of providing the service required more spectrum than another method, the economic calculus would result in assigning the spectrum to users of the former method (assuming the opportunity cost of the spectrum is positive). But if the consumer value of additional listening options exceeds the costs (including the spectrum opportunity costs) of providing those options, the economic approach would also advise allocating that additional spectrum. In the case of FM translators, it is our understanding that if the spectrum for FM translators is not used because of the lack of programming flexibility, the spectrum will simply lie fallow, unavailable for any other use. Thus, given the FCC's regulatory framework, the opportunity cost of this spectrum may be zero.

The approach to spectrum efficiency in the FM translator and the ATV proceedings seems consistent with the FCC's historical approach to the efficient use of the spectrum (Robinson, 1977). We do not mean to suggest that program diversity effects or the opportunity cost of spectrum use are not being considered by the agency in either if these two proceedings. However, even a cursory reading of the relevant agency discussions suggests that diversity concerns carry little weight. In the ATV proceeding, the FCC also noted that reallocating some of the useable ATV spectrum from non-broadcast use to ATV use would entail positive costs. However, the agency had no evidence that the costs of reallocation outweighed the benefits.

Marginal Station and Format Effects on Total Radio Listenership

To estimate the effects of the number of stations on program formats and total listenership, we estimated two equations. The first equation relates program diversity, as measured by the number of different formats[3], to the number of stations in the market. In this equation, we also control for market population, a series of demographic variables, and median household income in the market. Theoretical treatments of broadcast programming suggest that ceteris paribus, increases in the number of stations in the market will increase the number of formats offered (Steiner (1952), Spence and Owen (1977)).[4]

---

[3] The formats are easy listening, rock, big band, black, classical, ethnic (most of which are hispanic), religious, classical, news/talk, jazz, country, and other. The number of formats is not the only measure of program diversity that has been used in the literature, but theoretically it seems most appropriate. For a discussion and critique of these other approaches, see Haring (1975). Haring also estimates the effect of the number of stations on the number of formats. However, because he appears to exclude some relevant variables (such as demographics and average income), his estimate of the station-format relationship may be biased. Another potential source of bias is the possible endogeneity of the number of stations, a possibility for which we correct in our statistical analysis.

Other things equal (including the number of stations), the number of differing formats will depend upon differences in program tastes among the market population: The greater the dispersion of tastes, the greater the number of formats that might arise. Possible dispersion in tastes is proxied by the percentage of the market population over age 34, the percentage of the population that is black, and the percentage of the population that is Hispanic. We have no priors on the signs of these variables' coefficients.

The effect of any particular programming taste on the profitability of offering the formats demanded depends also on the absolute number of consumers sharing that taste. The greater the number of such potential listeners, the greater the potential advertising revenues, and the more likely is it that any particular format will prove profitable to offer. Thus we include market population as an additional independent variable.

Finally, we include median household income in the market. Other things equal, higher incomes may indicate the larger potential advertising revenues which would provide financial support for more formats.

---

[4] While the FCC determines the number of radio stations in each market, the number of stations might be considered an endogenous variable to the extent that the allowed number of radio stations is greater than the number operational. While a higher listenership may result from a greater number of operational stations (and therfore a greater number of formats), the number of operational stations (and therefore the number of formats) may depend upon listenership. We attempted to correct for this possibility by creating an instrument for the number of operational stations and the number of formats. The independent variables for the instruments include population, demographics, an estimate of the average station wage in each market, and the maximum number of possible FM stations (as determined by the FCC's table of allocations and as augmented by a number of recent FCC proceedings such as Docket 80-90). Because the FCC assigns new AM stations on a case-by-case basis, we were unable to include a measure of the maximum number of possible AM stations in each market.

The second equation assesses the determinants of the probability that a member of the market population will be listening to the radio at any point in time. The average proportion of the population listening to the radio during the day is hypothesized to be related to the number of formats, the demographic variables, and median household income. Other things equal, the greater the number of formats, the higher is the proportion of the population listening to radio likely to be. The demographic and income variables are assumed to be correlated with a "taste" for listening to radio. For example, if radio listening is a normal good, then higher incomes should lead to more radio listening.[5]

For the equation assessing the determinants of the number of formats, the coefficients of the number of stations, market population, and the percentage of blacks in the market are all positive and significant at better than the 5 percent level. The hispanic percentage is also positive at better than the 1 percent level of significance. The age and income variables are insignificant. While positive and statistically significant as suggested by theory, the effect of the number of stations on the number of formats is quantitatively small: Other things equal, a 10 percent increase in the number of stations leads to a 2.1 percent increase in the number of formats.[6]

---

[5] To estimate this equation, we use a weighted log-of-the odds model where the market weights vary with the sample size in each market used by Arbitron, Inc. We also note that we could have directly estimated the effect of the number of stations on the listenership probability without performing the intermediate step of estimating the number of stations on the number of formats. However, because population would then be an explanatory variable in the listenership equation and because it is highly correlated with the market sample size, such a weighted listenership equation contains substantial multicollinearity. To ameliorate this multicollinearity, we adopt the two equation approach described in the text. Because using this approach permits the number of stations to affect listenership only via the effect on the number of formats, we are likely to understate the effect of the number of stations on listenership. To the extent that the number of local radio stations is less than that required to reduce profits to a competitive level (because of FCC allocations policies), the introduction of additional stations may increase the rivalry for listeners and may create incentives for stations to devote some of their excess profits to increased programming expenditures. Higher programming expenditures may indicate a higher quality of radio service (e.g., compact discs instead of vinyl records, more entertaining or informative on-air personalities) that would attract additional listeners. For a discussion of this possibility with empirical support from television broadcasting, see Fournier (1985).

In the listenership equation, only the coefficient of the number of formats is statistically significant (at better than the 5 percent level). Our results suggest that other things equal, a 10 percent increase in the number of formats results in an average 1.9 percent increase in listenership.

Table 1 compares the effect of adding stations on listenership in a "small" and "large" market setting under two alternative assumptions regarding the percentages of blacks and hispanics in the population.[7] In both market settings, higher proportions of blacks and hispanics are associated with a greater listenership probability. For both market environments, a doubling of these proportions leads to an increase in listenership of between 3 and 4 percent. This in turn may indicate that these populations place a higher value on increases in radio service than other market segments.

For the small market case, an increase in the number of stations by 15 increases listenership by 6 to 7 percent (depending upon the demographic composition in the market). For the same absolute increase in the number of stations in the large market, the listenership probability increases by about 2 percent.

---

[6] There was virtually no difference between the equation that did not use an instrument for the number of stations and that which did (except as noted below).

[7] The effect of the number of stations on listenership (through the format effect) is statistically significant at the 10 percent level. In the format equation in which an instrument for the number of stations is not used, the magnitude of the effect is virtually identical but significant at the 5 percent level.

Table 2 illustrates the effect of increasing the number of formats on the percentage of the large market population listening to the radio.[8] For market populations with a higher proportion of blacks and hispanics, the listenership probabilities are 1 to 2 percent higher. An increase in the number of formats by 6 raises the listenership probability by 17.8 percent in the market with the lower proportion of blacks and hispanics and by 16.6 percent in the markets with a higher proportion of these groups.

On the one hand, we do not mean to suggest that in either market environment the additional 15 stations in Table 1 or the implied increase in the number of stations in Table 2 would be financially viable. Our purpose in this exercise is simply to illustrate the extent to which total listenership increases as the number of stations and the number of formats rises (thereby indicating that consumers place some positive value on that increase).

On the other hand, there is some reason to believe that the increased listenership (through the effect of the increase in the number of stations on the number of formats offered) may be understated. Low-power FM translators may find it more profitable to broadcast less popular programming than full-power stations by virtue of their lower power (and therefore lower cost) and by being able to locate the transmitter near smaller geographically-concentrated groups of listeners that place a high value on such programming. As a result, the number of formats may be more

---

[8] In this case there is little difference between the small and large market cases because population does not have a direct effect on listenership probabilities. The effect is indirect: The number of formats is dependent on the number of stations in the market which in turn depends upon the market population.

Table 1

Listernership and The Number of Stations

Small Market Case

(Population = 72.4 thousand; Medium Household Income = $28,582;
Population Percent over 34 = 10.2)

Number of Stations	Percent of the Population Listening to Radio 10% Black/5% Hispanic	 30%Black/15% Hispanic
5	11.9	12.4
10	12.3	12.8
15	12.5	13.0
20	12.7	13.1

Large Market Case

(Population = 2.4 million; Mediam Household income = $27,518);
Population Percent over 34 = 10.2)

Number of Stations	Percent of the Population Listening to Radio 10% Black/5% Hispanic	 30% Black/15% Hispanic
20	13.6	14.1
25	13.7	14.2
30	13.8	14.3
35	13.9	14.4

Table 2

Listenership and the Number of Formats
(Large Market Case)

Number of Formats	Percent of the Population Listening to Radio	
	10% Black/5% Hispanic	30% Black/15% Hispanic
5	14.0	14.4
7	15.0	15.4
9	15.8	16.1
11	16.5	16.8

responsive to the number of stations (including translators) than our statistical analysis suggests if the FCC were to grant translators programming flexibility.

Inframarginal Station Effects on Radio Listenership

We now turn to assessing the listenership effect of additional stations within a particular format. If all stations within a format offered programming equally attractive to consumers (in the sense that listeners in that format are indifferent among the stations within that format), then an additional station within that format would result in net consumer losses with no change in the number of listeners to that format.[9] Alternatively, an increase in the number of within-format stations may be valued by consumers because of the greater choice of listening options available. Consider a single classical music station that offered listeners only Mozart. If a second station entered the market and offered listeners Bach, total listenership of the format might in fact increase.[10] But even if total format listenership does not increase, consumers may nonetheless value additional stations within the format if additional stations permit greater catering to specialized tastes within the format. We discuss this possibility in the next section.

To empirically assess whether additional stations within a format increase listenership of the format, we related the share for each of the radio formats to the number of stations within the format. We also statistically controlled for the demographic composition of the market, median household income, the number of competing format choices, and the average "quality" of those competing choices. Our hypothesis suggests that ceteris paribus, the greater the number of stations in the format, the greater the format share. The demographic and income variables are included as proxies for format "tastes."

---

[9] This is clearly the assumption in Steiner (1952) as well as in Borenstein (1986).

[10] This hypothesis is similar to that made by Besen and Johnson (1982) with regard to some kinds of cable programming services.

The number of competing choices is included to control for the possibility of listener substitution across formats: Other things equal, the more format choices in the market and the greater the listener substitutability across formats, the less the share of any particular format. Similarly we control for the "quality" of these competing choices, measured by the average number of stations in the competing formats. Other things equal, the higher the quality of the competition, the smaller the share of the format. How important these competing choice variables will be depends upon the willingness of consumers to listen to various formats. For some formats, substitutability may be high while for others it may be small.

We first estimated this relationship for each format separately by ordinary least squares, the use of which requires the assumption that (in particular) the number of within-format stations, the number of competing formats, and the quality of the competing formats in the market are given or pre-determined.[11] For each of the formats but one (classical), the results were consistent with our expectations. The coefficient of the number of within-format stations was positive and statistically significant at better than the 1 percent level. The coefficient of the number of competing formats was negative, but significant in only two of the format equations while coefficient of the quality variable was negative and significant in three of the format equations. Of the demographic variables, the percentage of the market population that is black played the largest role, statistically significant in five of the format equations.

---

[11] Our analysis considered the listenership effects on the following formats: easy listening, rock, classical, country, ethnic, black, and news/talk formats.

One potentially serious problem with this analysis is that the number of within-format stations, the number of competing formats and the quality variable may all be endogenous rather than pre-determined. For example, the format share (given market population) is probably a good proxy for the total revenues generated by the format. Given station costs, more stations will tend to offer formats with the higher listenership. Thus, format share may depend on the number of stations within the format but the number of within-format stations may also depend on the format share. This kind of mutual dependence will bias our coefficient estimators. The coefficient on the number of stations will include not only the effect of the number of stations on increasing format share but also the effect of share on increasing the number of within-format stations. In the extreme, the causality may be the reverse of that assumed in our ordinary least squares analysis: If format listenership were predetermined (that is, it did not vary at all with the number of stations in the format), we would observe a positive relationship between format listenership and the number of within-format stations because a higher format listenership attracts more stations, not because more stations increase format listenership.

To correct for this potential problem, we created "proxies" for the number of within-format stations, the number of competing formats and the quality of the competition variable.[12] In effect, these proxies take into account the possibility of mutual dependence between format share and these three variables. Our results using the proxies contrasted sharply with the previous results: In no case were any of the variables (including the demographics) significantly related to the format listenership share. Thus we cannot reject the hypothesis that the number of stations within a format, the number of competing formats, and the quality of the competition have no effect on format listenership.[13] The contrast in results is also consistent with the hypothesis that (for example) the direction in causality runs from the format's share of listenership to the number of stations within the format.[14]

Consumer Indifference among Stations Within a Format

The results for the effects of the number of stations on total listenership and format listenership are consistent with the following conclusion: Spectrum efficiency from the standpoint of consumer welfare would best be attained (First Amendment considerations aside) by permitting at most only one station per format in each market. If consumers are indifferent among the stations within a format, permitting more than one station per format generates no additional consumer benefits but requires more resources in terms of station costs and the opportunity costs of the additional spectrum required. In this environment, new station assignments should be granted (if at all) only when the station license is conditioned on the offering of a format different from those already available in the market.

---

[12] More precisely, we developed instruments for each of the three right-hand side endogenous variables, using the same independent variables are those indicated in note .

[13] Thus, these results are consistent with those of Borenstein (1986) who finds some evidence that increases in the share of one station within a format comes at the expense of other stations in the format.

[14] Nonetheless, there are some reasons to pause before accepting these results. Most importantly, we find it puzzling that (for example) the percentage of blacks in the population has no effect on the listenership of black formats. In some cases, we did observe a high degree of multicollinearity among the instrumental variables and between these variables and the others in the analysis. While it is also possible that our selection of instruments could be improved, the regressions used to predict the endogenous variables were not obviously lacking in explanatory power, although some of the instrumental equations did have low R^2's. We are considering attempting a more complete specification of the instrumental regressions. In addition, we could improve the efficiency of our estimators by taking into account the likely non-zero covariances across the format regressions. Finally, not all of our sampled markets offered each of the formats considered, another factor we hope to account for in future research.

In this section we consider the possibility that listeners of a particular format are indifferent among stations offering that format. A finding that shares are equally distributed among stations within a format would be consistent with listener indifference.[15] If the programming of each station within a format is equally attractive to listeners, then the expected share of each station should tend to be the reciprocal of the number of stations offering the format. Uneven share distributions would be inconsistent with indifference.[16]

---

[15] However, equal shares could also arise if the tastes for within-format types of programs were uniformly distributed throughout the listening population. For example, if 50 percent of classical music listeners preferred Mozart and 50 percent preferred Beethoven, we might observe in a two-station format (one station specializing in Beethoven, the other in Mozart) an equal distribution of shares. But of course, deleting one station in the format may make at least one group of listeners worse off.

[16] It is possible but unlikely that an uneven share distribution could arise even if consumers are indifferent among the stations within the format. In particular, this could occur if after a listener has tuned to a station, the costs of then switching stations were high. Over time, the random selection of stations by new listeners would then result in a skewed share distribution. But it seems likely that most consumers listen to more than one station during the day, suggesting that switching costs are not substantial. Even

Even a cursory review of the within-format distribution of station shares reveals considerable dispersion. However, whether that dispersion is consistent with identical within-format tastes requires that we be able to compare the measured share dispersion to an hypothetical share dispersion when the expected share of each station within the format is simply one divided by the number of stations within the format.

Unfortunately, we are unaware of any valid statistical test that would enable us to make this comparison.[17] Thus we performed the following admittedly incomplete test. We first asked what is the probability of observing the highest station share within a format if the expected share is one over the number of stations in the format. A finding of a low probability of that high share occurring under the equal shares hypothesis could nonetheless be the result of sheer randomness. To partially account for this possibility, we calculated the probability of the second highest share occurring under the equal shares hypothesis after eliminating the first station (and its share of listeners) from the format. We repeated the exercise for the third and fourth highest ranked stations in the market when the number of remaining stations within the format was equal to or greater than two.[18]

---

for those consumers that listen to only one station, it seems implausible that the reason for such loyalty is the high cost of switching stations.

[17] The available statistical tests require that share of each station within a format be independent of the share of each other station. For any statistical test based upon all stations in a format, this requirement will not be fulfilled. If there are five stations in a format, and the share of four are known, then the share of the fifth is also known because the sum of the five shares must total 100 percent of the listenership in that format.

[18] We are grateful to Severin Borenstein for suggesting this conditional probability approach to us.

We performed this analysis for eight formats in the top ten radio markets, eliminating those formats in which the number of stations offering the format were less than three. There were forty-three market-specific formats examined. Table 3 provides some illustrative results. In most cases (not shown in Table 3), the probability of observing market shares equal to or greater than those of the top ranked stations under the equal shares assumption was less than one-tenth of one percent or less than one in a thousand. Only in one case (shown in Table 3) was the probability substantially higher. Thus, based on these results alone, one can reject the hypothesis that consumers are indifferent among stations within a format.

While we have not provided (and perhaps cannot provide) any evidence explaining the sources of these differences in station shares, a number of plausible explanations exist. For example, the observed share dispersion may reflect greater within-format specialization that is of value to consumers, in which case an increase in the number of stations may enhance specialization. It may also reflect the ability of some stations in a format to develop the programming formula (in terms of scheduling, on-air personalities, mix of music, news, and commercials) that is most attractive to listeners. Again, an increase in the number of stations may foster the search for the right "chemistry" by increasing the rivalry for listener patronage.

This is not to say that our results are completely convincing. One obvious criticism is that our analysis does not distinguish between AM and FM stations and had we controlled for this distinction, we would likely be unable to reject the equal share hypothesis. However, it is not at all obvious that this modification would dramatically alter our results. For example, virtually all rock stations in the top ten markets are on the FM

Table 3

Probability of Observing Actual Station Shares Under Assumption of Equal Shares Within A Format[19]

Market/ Format	Rank by Listener Share	Actual Share (%)	Predicted Share (%)	Probability of Observing Actual Share
A/Ethnic	1	44	25	Less than .001
	2	60	33	Less than .001
	3	60	50	Less than .05
B/Rock	1	16	9	Less than .001
	2	19	10	Less than .001
	3	21	11	Less than .001
	4	24	12	Less than .001
C/News-Talk	1	49	17	Less than .001
	2	65	20	Less than .001
	3	71	25	Less than .001
	4	55	33	Less than .004
D/Big Band	1	64	34	Less than .001
	2	90	50	Less than .001
E/Black	1	71	34	Less than .001
	2	51	50	.45

---

[19] In calculating the actual and predicted shares below the top ranked station, we deleted the higher-ranked stations from the format. For example, the actual share of the third-ranked station was calculated by deleting from the total listeners to the format the number of listeners for the two top-ranked stations. The predicted share of the third-ranked station under the equal shares assumption is calculated as:

l/(total within-format stations minus two) band and there is only one market (and only one station) in which a news/talk format does not appear on the AM band. Nonetheless, we could not statistically accept the equal shares hypothesis for these two formats in any market.[20]

A more serious critique is that our results might have been different had we started at the lower end of the market share spectrum or had we alternated between the two. These are possibilities among others we plan on pursuing, although we suspect that in the former case the results will be no different than those presented here and in the latter case, we will have to eliminate a substantial percentage of the stations before we find we cannot reject the equal shares hypothesis. We also intend to determine whether in fact an increase in the number of stations is related to greater dispersion.

Conclusion

Our results suggest that replacing the FCC's engineering oriented approach to spectrum efficiency with the economic approach would lead to a higher policy weight attached to the consumer benefits that may result from increasing the programming flexibility of translator licensees. Through its effect on the increase in the number of formats, an increase in the number of stations will increase the percentage of the population listening to radio. This in turn indicates that consumers place a positive value on the availability of additional formats. Moreover, our results suggest that the value of additional stations and formats may differ with the size and demographic composition of the market. While the listenership share of individual formats is unaffected by the number of radio stations offering the format, the dispersion in station listenership shares within a format suggests that consumers are not indifferent among the within-format stations.

---

[20] One could also argue that within both bands, not all stations are alike in that they differ across characteristics such as allowable power and (for AM stations), the ability to operate at night. We hope to incorporate these considerations in our future research, but given the wide dispersion of shares, we suspect that controlling for these characteristics would not alter our conclusions.

Thus use of the engineering approach to spectrum efficiency is likely to result in an allocation of radio spectrum and a use of that spectrum less preferred by consumers than that which would result from an economic approach. In the case of permitting FM translators more programming flexibility, our analysis suggests that the diversity benefits of such a policy have the potential for improving spectrum efficiency when efficiency is defined in terms of consumer welfare. These benefits derive from the effect of an increase in the number of stations on the offering of new formats, on the likely catering of stations to specialized tastes within a format, and on the likely within-format competition to find the most attractive programming formula.

To be sure, allowing more flexibility in spectrum use by translators may result in some displacement of full-power stations. And the FCC's engineering approach to spectrum efficiency has generated a reluctance on the part of the agency to increase the probability of such displacement by allowing such flexibility. The economic approach suggests that displacement even if it occurs may be welfare-increasing if the value (net of costs) listeners place on the translator programming exceeds that placed on the programming of the displaced station.[21] And if the characteristics of translators permit more formats to be offered for any given number of stations (the sum of full-power stations and translators), our results suggest significant increases in radio listenership.

---

[21] This would be more likely to be true if those listeners that preferred the programming of the full-power station to that of the translator found a close substitute in another full-power station within the same format. In addition, because translators operate at lower cost than full-power stations and because there could likely be many translators within any

We believe that at a minimum, our results indicate that consumers might be made better off if the FCC were to accord diversity-related consumer benefits a higher priority in its policy deliberations. However, as we noted at the outset of this paper, our analysis is incomplete in that we are unable to place a dollar value on these benefits and compare the value of these benefits to the costs of providing them. But explicitly or implicitly, the Commission will be rendering judgments about the significance of the dollar value of these benefits and costs when it finalizes its policy toward translators. Analyses such as the one presented here can reduce (but in a world of costly information, not eliminate) the scope of arbitrary and subjective policy judgments regarding the administrative allocation and use of the spectrum.

---

given market, new translators might arise to satisfy the preferences of these displaced listeners within a regulatory regime that permits translator licensees more programming flexibility.

References

Besen, Stanley, and Leland Johnson, An Economic Analysis of Mandatory Leased Access for Cable Television (Rand, 1982).

Borenstein, Severin, "Too Much Rock and Roll?: Business Stealing Effects in Radio Station Format Choice" (Paper presented at the Econometric Society, 1986).

__________ "On the Efficiency of Competitive Markets for Operating Licenses," Quarterly Journal of Economics (1988).

Coase, Ronald, "The Federal Communications Commission," Journal of Law and Economics (1959).

Federal Communications Commission, Notice of Inquiry in the Matter of Amendment of Part 74 of the Commission's Rules Concerning FM Translator Stations (1988a).

__________ Tentative Decision and Further Notice of Inquiry In the Matter of Advanced Television Systems and Their Impact on The Existing Television Broadcast Service (1988b).

Fournier, Gary, "Nonprice Competition and the Dissipation of Rents from Television Regulation," Southern Economic Journal (1985).

Haring, John, Competition, Regulation, and Performance In the Commercial Radio Broadcasting Industry, (Ph.D. Dissertation, Yale, 1975).

Kwerel, Evan, and Alex Felker, "Using Auctions to Select FCC Licenses," FCC Office of Plans and Policy Working Paper No. 16 (1985).

Mueller, Milton, "Property Rights in Radio Communication," (CATO Institute for Policy analysis, 1982).

Robinson, John, "Spectrum Allocation and Economic Factors in FCC Spectrum Management," IEEE Transactions on Electromagnetic Compatability (1977).

Spence, Michael, and Bruce Owen, "Television Programming, Monopolistic Competition, and Welfare," Quarterly Journal of Economics (1977).

Steiner, Peter, "Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting," Quarterly Journal of Economics (1952).

Webbink, Douglas, "Alternatives for Developing Spectrum Markets," (Paper delivered at the Conference on Market Alternatives to Regulations and Subsidies, 1981).

STATISTICAL APPENDIX

In this Appendix, we describe the econometric models used to depict the relationships discussed in the text among the number of radio stations, formats, and listenership. We first develop an equation relating the number of formats in a market to the number of stations. Then we present a model to estimate the relationship between the percentage of the market population listening to the radio and the number of formats. With these two equations, we are able to estimate the impact on listenership of increasing the number of stations. Last we develop an equation relating the number of listeners to a specific format to the number of stations in that format.

As discussed above, an increase in the number of radio stations may increase the number of programming formats offered, and increasing the number of formats may increase listenership. We can model the general relationship among the proportion of the market population listening to the radio, the number of stations and the number of formats as two equations; one relating the radio listenership to the number of formats and one relating the number of formats to the number of stations. These relationships can be depicted as follows:

$$P = F(FOR, X_1, v) \tag{1}$$

$$FOR = F(NS, POP, X_1, u) \tag{2}$$

where FOR equals the number of formats, NS equals the number of stations in a market, and P equals the proportion of population listening (called listenership below). POP is the population in the radio market. X_1 is a vector of other market characteristics that might affect listenership and the number of formats, and u and v are residuals.

We include POP in equation (2) because the number of formats in a given market may be influenced by population. One would expect a greater market population to increase the number of formats because large markets may have more people with diverse tastes and may generate sufficient advertising revenues to support differing formats. Consequently, stations offering more specialized formats in such markets may be more profitable than in markets with smaller populations.

We can depict the individual format listenership models as follows:

$$PF_i = F(NF_i, QF_i, FORF_i, X_1, w) \tag{3}$$

where PF_i equals the proportion of the radio audience in a market listening to format i (called market share below), and NF_i equals the number of stations within format i. QF_i is an indicator of the quality of the competition to format i: equalling the ratio of the number of stations in the competing formats to the number of these other formats in the market. $FORF_i$ is the number of formats other than format i in the observation market; it represents the amount of competition to the format within the market.

The X_1 vector consists of variables that may affect the distribution of programming tastes across given populations and their use of radio as an entertainment medium. As such, X_1 may affect all three of our dependent variables, P, FOR, and PF_i, although we do not have any priors on whether the coefficients of the components of X_1 in these equations should be positive or negative. The components of X_1 are median household income (Income), the population proportion of Blacks (Black) in the market, the population proportion of Hispanics (Hispanics), and the proportion of the population over 34 years old (Age).

For the sample, we examined the 259 Arbitron radio markets for which industry data are collected. To insure that our data points closely correspond to real markets, we confined our sample to the 115 markets where the local stations have a market share of over 85 per cent. We used data for the spring of 1987.[1] For the formats, we aggregated the 42 different types of station formats categorized by Broadcast Investment Analysts, Inc. into 11 more inclusive formats. The aggregated formats are as follows: Easy Listening (Easy), Rock, Country, Big Band, Black, Religious, Classical music (Classical), Diverse, Ethnic (the bulk being Hispanic), News (consisting of not only news stations but also talk, sports, and farm stations) and Jazz.[2]

---

[1] Most data are from Broadcast Investment Analysts, Inc.(1988). We express our appreciation to Tom Buono and Gregg Welsh of Broadcast Investment Analysts for their permission to use the data. Listenership data are from the market-by-market ratings publications of Arbitron.

[2] We assigned the Broadcast Investment Analysts, Inc. formats to our eleven aggregated ones as follows:

Easy - Adult Contemporary, Beautiful Music, Contemporary, Easy, Middle-of-the-Road, New Age, Niche 29, and Soft Adult Contemporary;

Rock - Album Oriented Rock, Contemporary Hit Radio, Classic Hits, Classic Rock, Oldies, Progressive, Rock, Soft Rock, and Top 40;

Country - Country;

Big Band - Big Band, and Nostalgia;

Black - Black, Black Oldies, and Urban;

Religious - Christian Contemporary, Christian, Christian Rock, Gospel, Inspirational, Religious Music, and Religion;

Classical - Classical;

Since it is likely that (for example) the absolute effect of the number of stations on format is dependent on the size of the population in the market, we estimate equations (1), (2), and (3) in double log form:

$$\begin{aligned} \ln FOR &= \ln b_0 + b_1 \ln NS + b2 \ln Pop + b_3 \ln Income \\ &\quad + b_4 \ln Black + b_5 \ln Hispanic + b_6 \ln Age \\ &\quad + u \qquad (1a) \end{aligned}$$

$$\begin{aligned} \ln [P/(1 - P)] &= \ln a_0 + a_1 \ln FOR + a_3 \ln Income + a_4 \ln Black \\ &\quad + a_5 \ln Hispanic + a_6 \ln Age + v \qquad (2a) \end{aligned}$$

$$\begin{aligned} \ln [PF_i/(1 - PF_i)] &= \ln c_0 + c_1 \ln NF_i + c_2 \ln QF_i + c_3 \ln FORF_i \\ &\quad + c_4 \ln Income + c_5 \ln Black + c_6 \ln Hispanic \\ &\quad + a_6 \ln Age + w \qquad (3a) \end{aligned}$$

In the listenership and format share equations (2a and 3a), we transform P (and PF_i) to represent the log of the odds in order to constrain the predicted P and PF_i to the zero-one interval.[3] We therefore use the minimum logit chi square method in estimating equations (2a and 3a).[4] Since P and P/(1 - P) are proportions, they have a range of from zero to one and P has a binomial distribution with a variance of P(1 - P), and ln[P/ (1 - P)] has a variance of $1/[n_k(1 - P)P]$ where n_k is the number of people in the Arbitron sample for market "k." Thus the error term v is heteroskedastic. To correct for this problem, we use a weighted least squares technique with the following weights $[n_k(1 - P)P]^{1/2}$ for

---

Ethnic - Ethnic, Hawaiian, Japanese, Polka, Portuguese, and Spanish;

News - Farm, News/Talk, and Sports;

Jazz - Jazz;

Diverse - Diverse.

[3] For brevity's sake, we use only P in the discussion on the log of the odds. In this paragraph, what applies to P also applies to PF_i.

[4] A discussion of this method can be found in Maddala (1983), p. 28-32.

the total listenership equation and $[n_k (1 - PF_i)PF_i]^{1/2}$ for the format share equations.[5]

In order to evaluate the impact of NS on P directly, it would seem logical to substitute the FOR equation into the total listenership logit equation. However, this creates the potential for a serious multicollinearity problem. Substitution of (1a) into (2a) results in ln POP being an additional weighted independent variable which is highly correlated with other weighted variables in (2a) because of the correlation (0.9008) between ln POP and the Arbitron sample size n_k. Such multicollinearity can lead to large variances for the parameter estimators thereby reducing the precision of our estimates. Thus it increases the probability of accepting the hypothesis that the coefficients are not significantly different from zero when in fact the null hypothesis should be rejected.[6] In fact, the correlation coefficients between the weighted ln POP and three of the other weighted variables (ln NS, ln Income, and ln Age) are 0.9935, 0.9974, and -0.9511 respectively.[7] Consequently, rather than substitute equation (1a) into equation (2a) and estimate a single equation, we estimate each equation and evaluate the impact of NS on P by making that substitution after the two equations have been estimated.

One could argue that since POP is implicit in the FOR equation, it still has not been eliminated as a source of multicollinearity when we use FOR only in the P equation. However, there are several other variables in the format equation, and the impact of population on the number of formats, while statistically significant, is not large (the regression coefficient being about 0.10). Thus its contribution to any multicollinearity problem through its effect on FOR is likely small. The multicollinearity is probably less of a problem in the FOR equation because we do not use the weighted estimation technique. For instance while the weighted ln NS and ln POP variables have a correlation coefficient of (0.9935), the correlation between the unweighted variables, while high (0.9122), is lower than the weighted versions of the variables.

Thus estimating the FOR equation for the relationship between the number of formats and the number of stations may enable us to obtain more precise estimates of the effect of station numbers on listenership. The method used here is one of the solutions to the multicollinearity problem recommended by Kennedy. In particular, he suggests that "formaliz[ing the] relationships among regressors" in the same manner as described above may reduce the impact of multicollinearity.[8]

---

[5] The basis for using these weights can be found in Maddala (1983), p. 28-32. We express our appreciation to Arbitron Inc. for providing us with the market-by-market sample data.

[6] See H. Theil (1971), p. 147-155.

[7] There are also high correlation coefficients between some of the other variables, but they are all required in the listenership equation for theoretical reasons. To exclude these relevant variables would lead to biased estimators.

The major estimation problem with these equations is that some of the independent variables may not be predetermined. In particular, NS in equation (1a) and FOR in (2a) along with NF_i, QF_i, and $FORF_i$ in (3a) may be endogenous.[9] To correct for this problem, we use an instrumental variables technique to estimate the three equations. The instruments are found by regressing the log of each of the five endogenous variables on the logs of a group of related variables. These variables are Pop, Income, Black, Hispanic, and Age which are already in the above equations plus two other variables, radio industry wages,[10] and the number of FM assignments for each market.[11]

Table A1 reports the results for the estimation of equations (1a) and

---

[8] P. Kennedy (1985), p. 151.

[9] FOR in equation 2a may still be endogenous; while P does not appear in equation 1a, it still may indirectly impact on FOR by way of its influence on NS, the number of stations.

[10] This variable is derived from the accounting data on radio stations compiled by the National Association of Broadcasters (1986). By making a plausible assumption about the number of hours worked by part-timers, we can use radio station accounting data to calculate the average wage for the radio station workers in each format. For each observation market, we compute a weighted average of the wages paid by the various formats present. Then, making use of the data on station wages in different population size classes, we multiply the weighted average for each market by the ratio of the computed wages in the markets of the same size class to the wages in a median size market class.

[11] These assignments can be found in the Broadcasting & Cablecasting Yearbook 1987 on p. B431-B440. In addition to these assignments, we include in each market area total the number of new FM assignments resulting from the FCC Docket 80-90 proceedings. This number can be found in Broadcast Investment Analysts, Inc. (1988).

(2a). Our estimating method takes into account the possibility of correlation between the residuals of the two equations by using the seemingly unrelated equations (SURE) method. With this technique, the two equations are estimated jointly (the variables in the listenership ratio equation being properly weighted as in a single equation minimum logit chi square estimate.)[12]

In the format equation, all the variables except Income and Age are significantly different from zero at the 5 percent level at least. The positive coefficient for NS is as predicted by theory; an increase in the number of formats is associated with a greater number of stations. The coefficient being considerably less than one supports the contention of Steiner and others on the tendency of advertiser supported radio to weight its offerings toward "common-denominator" programming.

The positive coefficients for the proportion of Blacks and Hispanics suggests that the programming tastes of these groups contribute to format diversity. The insignificant results for Income and Age are not particularly alarming because they are not inconsistent with economic theory. Nothing in the literature states that differences in age and income would necessarily lead to more diversified tastes in radio programming, other things equal.

In the listenership equation, the coefficient of the number of formats is positive and significantly different from zero at the 5 percent level. The proportion of Hispanics in the population is also significantly different from zero but only at the 10 percent level. The other demographic variables, however, are not significantly different from zero on a two tail test. (Since economic theory makes no prediction as to the direction of either coefficient, we rely on the two tail test.) Again these insignificant results are not particularly alarming because economic theory only predicts a possible effect for these variables.

As a whole, the format and total listenership equations are statistically significant; the log likelihood ratio for the equation system is 96.73 which is significant at the 1 percent level.

The results for the format share equations are less conclusive. When we estimate these equations by an OLS technique assuming that NF_i, QF_i, and $FORF_i$ are exogenous and using the log of the odds or minimum logit chi square methodology, the number of stations coefficient is usually significantly different from zero, and some of the other variables are significantly different from zero with theoretically plausible signs. With the asymptotically unbiased instrumental variables technique, however, the number of stations within the formats is not significantly different from zero for any of the six format share equations. Furthermore very few of the other variables are significant with this technique. It is possible that the number of stations does not influence the number of listeners in a given format, but the possibility that estimation problems are responsible for the results can not be ruled out.

---

[12] For a discussion of SURE, see J. Johnston, (1984), p. 330-341.

These results leave us in a quandary; on one hand, given the simultaneity between many of our variables, the OLS technique is very likely to be biased. On the other hand, while an instrumental variable technique is asymptotically unbiased, multicollinearity between some of the instruments used in these market share equations may lead to upwardly biased parameter variance estimates. For instance the correlation coefficient between the instruments for NF_i and $FORF_i$ is 0.9654 for the rock format. Therefore we can make no definite conclusions except that our results do not contradict the null hypothesis that market share of a format is not related to the number of within format stations.

The implications of our findings for FCC policies are as follows. The fact that increasing the number of formats increases listenership implies that the marginal listener values additional alternative types of radio programming. Thus consumers may benefit if given radio markets had more formats.

Our model also suggests that by increasing the number of stations, a more flexible policy toward translators will indirectly lead to increased listenership because of the effect of the number of stations on the number of formats. By multiplying the coefficient for NS in the format equation by the coefficient for FOR in the listenership equation, we can calculate the impact of a change in the number of stations on the proportion of the population listening to the radio. For our instrumental variables SURE estimation model, this product is 0.208 times 0.233 or 0.048. An asymptotically normal variance can be calculated for this product as:[13]

$$\mathrm{Var}\,(b_1{*}a_1) = b_1^2\,\mathrm{Var}\,(a_1) + a_1^2\,\mathrm{Var}\,(b_1) + 2\,b_1{*}a_1\,\mathrm{Cov}\,(b_1, a_1).$$

For our model, the square root of this variance equals 0.0306. This gives a t value of 1.59; thus the product of the two coefficients is significantly greater than zero at the 10 percent level.

While these calculations do not indicate that increasing the number of stations will greatly enlarge listenership, they do imply that the consumers would value more listening options -- especially stations with different formats.

The inconclusive results for the format share equations combined with the strong relationship between the total listenership and the number of formats suggests that an FCC policy of only granting station licenses to firms willing to set up new formats might be the most efficient in terms of consumer welfare. More stations within formats does not seem to increase format audience while increasing formats increases the total radio audience. Thus, adding stations to already existing format will not increase audiences as much as adding stations with new formats.

---

[13] G. C. Chow (1983), p. 182-184. In order for this formula to apply, the estimators, a_1 and b_1, must be asymptotically consistent. This assumption is fulfilled by our instrumental variables technique.

The problem with such a policy is that it assumes that all firms have the same ability to serve the listeners in a given format. This implies that on the average stations within given format would have roughly the same market share. If consumers were indifferent between radio stations within a format and any firm licensed by the FCC could on the average serve a format as well as any other, we would expect to see the within format station market shares distributed binomially about the mean share.

To test this hypothesis we examine the actual market shares within formats for the top radio ten markets. If the market share is distributed binomially, its expected variance is as follows:

$$\text{Variance}_{ik} = n_k^{-1} (\text{Average Market Share}_i) (1 - \text{Average Market Share}_i)$$

for format i in market k:[14] the standard deviation being the square root of this variance. Since in large samples the binomial distributions assumes a normal shape, one can calculate the probability that any given market shares would occur under a regime of evenly distributed shares. We find this probability by first dividing the difference between the given share and the average by the calculated binomial standard deviation. Then we compare the actual occurrence of this market share with the Gaussian (normal) probability of this market share occurring.

After calculating the difference between the actual occurrence and the expected probability for the largest firm in the format; we repeat the exercise for the second largest firm in the format. In calculating the relevant statistics for the second largest share, we deduct one from the station sample size and subtract the share of the largest firm from the total market size. We did this exercise for 43 market-specific formats among the ten largest markets.

The probability of observing the actual market shares in an equal market share regime is usually about one thousand to one. Thus we usually reject the hypothesis that their distributions were normal (Gaussian) with the average market share as the mode and median. This implies that a considerable amount of within format diversity exists, and thus consumers value the ability to choose different variants of the same format (Oldies v. Heavy Metal).

These results suggest that stations within the same format may be quite different: serving a quite different functions or a different set of consumers than another. Consequently even though adding a new station to an established format may not increase format listenership, it may increase consumer welfare by increasing within format diversity and widening consumer choice.

---

[14] See Wonnacott and Wonnacott (1977), p.82-88 and 171-174.

Table A1
The Relationship between the Proportion of Population Listening to the Radio (Listenership Ratio) and the Number of Formats and the Relationship between the Number of Formats and the Number of Stations[1]

Variables	Coefficients for the Format Model	Coefficients for the Listenership Model
Constant	0.287	-0.844
Number of Stations	0.208 (2.16)	________
Number of Formats	________	0.233 (2.28)
Population	0.094 (2.38)	________
Income	-0.013 (-0.12)	-0.101 (-1.02)
Black	0.024 (2.45)	0.005 (0.45)
Spanish	0.040 (3.49)	0.018 (1.38)
Age	-0.047 (-0.27)	0.107 (0.70)
Log Likelihood Ratio	96.73	96.73

---

[1] "t-statistics" are in parentheses.

Appendix References

Broadcast Publications, Inc., Broadcasting & Cablecasting Yearbook 1987, Washington D. C.: Broadcast Publications, Inc.

Broadcast Investment Analysts, Inc., Investing in Radio, 1988, Washington D. C.: Broadcast Investment Analysts, Inc. 1988.

G. C. Chow, Econometrics, New York: McGraw-Hill Book Company, 1983.

J. Johnston, Econometric Methods, Third Edition, New York: McGraw-Hill Book Company, 1984

G. S. Maddala, Limited Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press, 1983.

P. Kennedy, A Guide to Econometrics, Cambridge, MA.: MIT Press, 1985.

National Association of Broadcasters, 1986 Radio Financial Report, Washington D. C. : National Association of Broadcasters. 1986.

H. Theil, Principles of Econometrics, New York: John Wiley, Inc., 1971.

T. H. Wonnacott and R. J. Wonnacott, Introductory Statistics, New York: John Wiley & Sons, 1977.

PLEASE DO NOT
CIRCULATE OR QUOTE

TRIGGER PRICE REGULATION*

David J. Salant

Economics & Statistics Department
GTE Laboratories Incorporated
Waltham, MA 02254

Department of Economics
Virginia Polytechnic Institute
Blacksburg, VA 24061

Glenn A. Woroch

Economics & Statistics Department
GTE Laboratories Incorporated
Waltham, MA 02254

July 1988

---

* We are especially grateful to Bob Rosenthal for detailed comments. We have also benefited from discussions with Jim Friedman and Roy Radner, and from comments supplied by Bill Taylor. The views expressed in this paper are those of the authors and should not be attributed to GTE Corporation or any of its subsidiaries.

ABSTRACT

Private enterprise natural monopoly must surrender some control of investment decisions to a rate setting regulatory authority acting as an agent of consumers. Additions to capital stock generate adjustment costs which both favors gradual accumulation and represents sunk costs. Once sunk costs have been incurred the regulator has no incentive to allow rates which enable the firm to recover its investment.

We begin by solving the planning problem of maximizing discounted consumer surplus while ensuring the firm breaks even. The result is a dynamic Ramsey-Boiteux pricing rule together with an efficient investment path approaching the optimal steady-state capital stock. The attempt to support this solution as a strategic equilibrium between the firm and regulator results in a "prisoners' dilemma."

To investigate equilibrium outcomes we formulate the relationship as a time-dependent supergame, and examine two extremes of the degree of commitment in the resulting game. First, we allow both parties to select at the outset a course of action for the remainder of time. Starting from a steady state, we find that a unique open-loop equilibrium exists in which price exactly covers operating cost and no new investment occurs.

Next we express the opposite case of no commitment using trigger-type history-dependent strategies. An equilibrium exists in which the regulator follows the efficient price path as long as the firm implements the efficient investment plan; if any deviation occurs, the regulator cuts price down to operating expense. Likewise the firm invests at the efficient level provided the regulator allows an adequate return, but ceases investing the moment profit dips below the agreed level. For a sufficiently large discount factor and depreciation rate, these trigger strategies support any steady-state optimum. Surprisingly, trigger strategies out-perform the open-loop outcome in a wide range of cases verifying that full commitment is inadequate to overcome the investment disincentives caused by strategic behavior.

A linear-quadratic example is worked out to demonstrate these steady-state results, and also to explore the effectiveness of trigger strategies along the dynamic path. It also allows us to perform simple comparative statics exercises. We find that the ability of trigger price strategies to support cooperative outcomes depends on values of initial and steady-state capacities.

KEY WORDS: price ceiling regulation, time-dependent supergame, trigger strategies

1. INTRODUCTION

A long-term relationship develops between a public utility and its customers when each is required to commit resources for an extended period of time. The firm undertakes an investment only when it expects an adequate return over the lifetime of the project which, in turn, depends upon regulatory policy toward price and entry. Similarly utility customers require assurances of an acceptable level of availability and quality of service before relinguishing rights of access to public thoroughfares or making complementary investments of their own. Goldberg (1976) viewed the principal duty of a regulatory agency to be the administration of the implicit "regulatory contract" which protects the respective rights to serve and be served.

The effectiveness of an agreement to sustain efficient investment hinges on the incentives for all parties to adhere to its terms *including the regulatory agency* when it acts on the behalf of consumers. As Williamson (1975) observed, whenever there is the need to invest in relationship-specific assets the investor becomes vulnerable to "opportunism" by a trading partner who possesses bargaining power.[1] Essentially, once an expenditure becomes sunk, only the avoidable cost associated with continuing the relationship governs future behavior. In the absence of binding contracts, any promises to provide returns on sunk investments are rendered worthless.

Upon installing highly specialized, highly immobile plant and equipment, a public utility's ability to minimize the impact on revenues caused by abrupt changes in tariff policy becomes severely limited. The regulatory agency can cut rates arbitrarily close to avoidable cost, thereby eliminating any return in excess of operating expenses without fear that service will be discontinued in the immediate future. As a result, a regulator can obtain much more favorable rates for its clients, at least until it comes time to replace the existing capital.[2] In practice such behavior takes its most direct form when a "imprudent" plant or unsuccessful R & D venture are disallowed from the rate base. More subtle but no less insidious instances of opportunism occur when an agency engages in "regulatory lag" or excessively slow depreciation.

The institutional setting surrounding most regulatory agencies does little to discourage opportunistic behavior. An agency must pay attention to the wishes of its overseers and to the dictates of precedent. Nevertheless, it retains a great deal of leeway with regard to the interpretation and implementation of its mandate. Furthermore, each electoral cycle brings a wave of political appointments to the agency who are only loosely bound to follow policies of their predecessors. The new staff members base their decisions on their own preferences and information.

Nor is the firm likely to have incentives to follow through on its proposed pricing and investment plans. For instance, following the initial startup, the firm's capacity will typically be quite small which will, in turn, justify relatively high prices and high returns for the firm. After reaping returns in excess of the market rate during early years, the firm may be able to escape the compensating losses by curtailing plans for capital expansion (and possibly replacement) and allowing service quality to deteriorate.

When each party behaves opportunistically toward the other's irreversible actions, rates of return are bound to fall off, resulting in an across-the-board disincentives to invest. Without the help of a third-party to police deviant behavior, and as long as play ends with some final period, the firm-regulator relationship is destined to suffer the inefficiency of the notorious "prisoners' dilemma."

Fortunately we can draw on an important lesson from the study of repeated games that demonstrates how repetition of the relationship may recapture the mutual gains available to the firm and its customers. The longer the relationship lasts, the greater the potential losses, and hence, the greater the benefits of voluntarily cooperating. The parties can do so by credibly threatening to revert to non-cooperative behavior indefinitely should either side pursue short-run gain. The emergence of such cooperation in the firm-regulator relationship is the principal goal of this paper.

To do so we model the control of natural monopoly as an infinitely-repeated interaction between a private enterprise and a regulatory authority that seeks maximum consumer welfare. Starting with an initial stock of capital, the utility invests in durable plant and equipment needed to provide a single service. Additions to the stock generate an adjustment cost which favor gradual accumulation. Furthermore, all capacity costs are assumed to be sunk due to physical immobility, specialization of function, and the high cost of transfering ownership. This depiction of the capital investment process is rather stylized but should serve to illustrate the tendencies caused by sunk costs. We are confident our conclusions continue to hold for any investment which yields returns spread over an extended period -- such as R & D aimed at cost-reducing innovations.

In order to establish a benchmark, we calculate the planning solution that maximizes consumer welfare over time while ensuring the firm's solvency. The result is a dynamic version of the Ramsey-Boiteux pricing rule together with an efficient investment path approaching the optimal steady-state capital stock. The remainder of the paper is concerned with how closely this solution can be approximated by a non-cooperative equilibrium.

We begin with the case in which both the firm and the regulator can commit at the outset to price and investment plans over an indefinite future, i.e., open-loop strategies. In each period the regulator sets a price ceiling that the actual price must not exceed. Quantity is bounded above by installed capacity as well as by realized demand. Immediately we see that the optimum cannot be supported as an equilibrium since both the firm and the regulator have an incentive to deviate from the required price-investment path. The regulator boosts consumer surplus by cutting price once capacity is installed, and the firm responds by ceasing all investment and curtailing production the moment that its market value becomes negative. A unique equilibrium exists in which price just covers operating cost and no investment takes place: strategic behavior reduces the outcome to a sequence of single-period equilibria.

Next, we consider the equilibrium when the firm and regulator can base their actions on the entire history of play up to that time. One type of such closed-loop strategies are those which have "trigger property" in which all players "agree" to follow a certain course of action unless one or more "defects", at which time they all revert to another set of specified actions for the remainder of the game. We focus on a trigger price strategy wherein the regulator offers the firm an above-market rate of return provided it conforms to the efficient investment plan, but if it deviates the regulator credibly threatens to cut price so as to just cover operating expenses. Symmetrically, the firm maintains the optimal capital stock until the return falls below a prescribed level, and at that instant it halts any further investment, allowing the capital stock to depreciate, and producing to just break even.

For a sufficiently large discount factor, these trigger strategies support any steady-state optimum as a subgame perfect equilibrium outcome. It is also shown that the lower the discount and depreciation rates, the closer the optimum can be approximated as an equilibrium. Somewhat surprisingly, we find a large class of cases in which the closed-loop equilibrium outperforms the open-loop outcome. Thus, the commitment usually attributed to open-loop strategies does not allow the regulator to overcome the investment disincentives of its opportunism. In addition to their ability to support efficient outcomes, trigger strategies have the advantage that they are fairly simple to express.

Finally a linear-quadratic example is worked out to demonstrate these steady-state results, and also to explore the effectiveness of trigger strategies along the dynamic path. We find that the ability of trigger price strategies to support cooperative outcomes depends on values of initial and steady-state capacities.

The next section sets out the assumptions of the model and characterizes the planning solution. The game played by the firm and regulator is described in Section 3. In Section 4 open-loop equilibria are studied when the system starts out in a steady state. Section 5 introduces closed-loop strategies, in particular the trigger price rule, and characterizes the steady-state equilibrium that they support. We then derive both steady-state and dynamic equilibria using a linear-quadratic example and perform some simple comparative statics exercises.

2. THE MODEL

Production and consumption occur at an infinite number of discrete points in time. The firm produces a single, perishable product with a well-known technology. Except for the accumulation equation, demand and cost conditions are separable across time. Several natural extensions of this simple setup are mentioned in the last section of the paper.

2.1 Demand and Cost Conditions

Let p_t and q_t be the *actual* price and quantity in period t. Output is constrained by the amount of installed capacity $\hat{q}_t$, and the price level is bounded above by a price cap $\hat{p}_t$. The demand curve is stationary and given by Q(p) from which we derive the inverse demand P(q) and the revenue function, R(q) = P(q)q. Often we will require a measure of consumer surplus taken net of an arbitrary outlay:

$$S(p,q) = \int_0^q P(x)dx - pq. \tag{1}$$

When the price charged exactly equals the demand price (i.e., $p = P(q)$), this becomes the usual Marshallian consumer surplus.

Production requires operating expenses and investment outlays. For simplicity we assume a constant unit variable cost of $v \geq 0$.[3] The cost of investing I in any period is given by the increasing, convex function C(I) where C(0) = 0.

Initial capacity is given by $k_0 \geq 0$ and it proceeds to depreciate at a geometric rate δ. Thus, if gross investment in period t is I_t, next period's capacity is just $\hat{q}_t = I_t + \gamma\hat{q}_{t-1}$, where $\gamma = 1 - \delta$ is the rate at which capital survives. Since investment is assumed to be irreversible (i.e., $I_t \geq 0$), depreciation occurs at a maximum rate of δ.

2.2 *The Objective Functions*

There is a constant market interest rate of r, and both the firm and the regulator discount the future using the factor $\rho = 1/(1+r)$. The regulator maximizes discounted consumer surplus given by:

$$\mathcal{S} = \sum_{1}^{\infty} \rho^{t} S(p_t, q_t) \tag{2}$$

Owners of the firm are solely interested in the returns their investment will yield. In the short run, the cash flow of the firm is just $\pi_t = (p_t - v)q_t - C(I_t)$ where $q_t \le \hat{q}_t$. In the long run they evaluate the performance of the firm according to discounted profits:

$$\prod = \sum_{1}^{\infty} \rho^{t} \Big((p_t - v)q_t - C(\hat{q}_t - \gamma\hat{q}_{t-1}) \Big) \tag{3}$$

where we have made the substitution $I_t = \hat{q}_t - \gamma\hat{q}_{t-1}$. The firm's problem, if it anticipates the price sequence $\{\hat{p}_1, \hat{p}_2, \ldots\}$, then reduces to a choice of a nonnegative sequence of capital stocks which satisfies the constraint $\hat{q}_t \ge \gamma\hat{q}_{t-1}$ for each t and that maximizes Π.

2.3 *The Planning Problem*

Here we characterize the second-best price-investment path subject to the constraint that the firm earns a given level of profit Π^0. A solution -- denoted by a star * -- consists of a sequence of capacities, prices and quantities $\{\hat{q}_t^*, p_t^*, q_t^*\}$ which maximize $\mathcal{S}$ subject to $\Pi \ge \Pi^0$, $p_t = P(q_t)$ and $q_t \le \hat{q}_t$, and $\hat{q}_t \ge \gamma\hat{q}_{t-1}$ for each t given an initial capacity k_0.

Some explanation of this setup is warranted. First observe that the regulator's objective places no explicit weight on the firm's profit. While this differs from many formulations of second-best problems, it does not sacrifice any generality. The firm will voluntarily enter into a relationship only when it anticipates it can do at least as well as the next best venture. The regulator is therefore constrained to provide the firm with at least this level of profit Π^0, thereby introducing an implicit weight on profits into its objective in the form of a Lagrangean multiplier.

That is not to say the firm's *short-run profit* π_t will never dip negative; in the early periods the high cost of rapid accummulation may very well swamp the firm's revenues. In fact, the market value of the firm Π^t -- the discounted stream of profits from period t onward -- could at some time become negative though it must ultimately remain nonnegative. As we will see, in order to prevent the firm from exiting the industry when $\Pi^t < 0$, or to prevent the regulator from reneging on its tariff policy when $\Pi^t > 0$, some form of "binding contract" is needed to sustain this planning solution.

Before turning to a strategic formulation, however, we will solve this problem. This is done by finding a singularity of the Lagrangean:

$$\mathcal{L} = \sum_{t=1}^{\infty} \rho^t \left\{ S((P(q_t), q_t) + \lambda\left[R(q_t) - vq_t - C(\hat{q}_t - \gamma\hat{q}_{t-1})\right] + \mu_t\left(q_t - \hat{q}_t\right)\right\} \tag{4}$$

Variants of this welfare problem have been solved on several occasions (e.g., Brock and Dechert (1985)). When there is no excess capacity (i.e., $q_t = \hat{q}_t$, or equivalently $\mu_t > 0$),[4] the first-order conditions reduce to a period-by-period, inverse-elasticity rule:

$$\frac{P_t - MC_t}{P_t} = \frac{\eta}{\varepsilon_t} \tag{5}$$

where, as usual, $\varepsilon_t \equiv -P(q_t)/P'(q_t)q_t$ is the price elasticity of demand, and $\eta = (\lambda-1)/\lambda \in (0,1)$ is the "Ramsey number". (The value of λ, and hence η, derives from the firm's budget constraint Π^0.) Here MC_t is a *dynamic* marginal cost:

$$MC_t = v + C'(I_t) - \gamma\rho C'(I_{t+1}) \tag{6}$$

The first two terms in (6) are simply direct operating and investment costs; the last term represents the savings in the form of reduced investment in the following period afforded by an additional unit of current investment, after properly discounting by ρ and deducting for depreciation with γ.

Finally when the system starts out with abundant capacity, the planning solution calls for no investment until capacity depreciates down to the steady state. In the meantime, price is set at operating cost (i.e., $p_t = \hat{p}_t = v$) and production exactly satisfies realized demand (i.e., $q_t = Q(p_t) < \hat{q}_t$).

2.4 The Optimal Steady-State Solution

The system is in *steady state* when investment exactly replaces depreciated capital: $I_t = \delta\hat{q}_t$. To indicate steady-state outcomes, we drop the time subscript and place a bar over the variables: $\hat{q}_t = \bar{q}$, $p_t = \bar{p} = P(\bar{q})$, etc. The *optimal* choice of steady-state capacity solves a simple static problem: maximize $S(P(q),q)$ subject to $[P(q) - v]q - C(\delta q) \geq 0$. The resulting solution $\bar{q}^*$ is related to δ from the static Ramsey-Boiteux rule that follows from (5):

$$(\bar{p}^* - \overline{MC}^*)/\bar{p}^* = \bar{\eta}^*/\bar{\epsilon}^* \tag{7}$$

where the term $\overline{MC}^* = v + [(r + \delta)/(1 + r)]C'(\delta\bar{q}^*)$ is the marginal cost of adjusting to and then sustaining a higher level of output and capacity.

3. THE STRATEGIC APPROACH

Up to this point, the firm obediently executes the investment plan proposed by the planner. Difficulties arise when a regulatory agency is entrusted with controlling a privately owned and managed enterprise. Naturally, the firm's managers prefer higher prices and a smaller capital stock than that dictated by the planning solution. A principal goal of this paper is to inquire whether it is possible to sustain the planning solution when the regulator controls price and the firm controls investment (and production).

We endow both the firm and the regulator with perfect recall and perfect foresight when selecting their strategies. At the beginning of each period the firm and regulator simultaneously choose a capacity and a price cap, respectively. The firm can elect not to enter the market by setting its capacity at zero for each period. The complete extensive form of the game is given Figure 1.

Given a price cap and a capacity constraint, the final price and quantity remain to be determined. As in the classical bilateral monopoly problem an indeterminancy exists: an entire range of price-quantity pairs satisfy the constraints each yielding different payoffs to the two parties. A popular compromise in such situations is to settle on the Nash bargaining solution. Although alternatives are considered, we focus on a rule that delegates the choice of final price and quantity exclusively to the firm. This rule, we believe, matches closes to the actual way in which price and quantity are determined in practice.[5]

Let the rule which determines price and quantity for each price cap and capacity be given by $p(\hat{p}_t,\hat{q}_t)$ and $q(\hat{p}_t,\hat{q}_t)$. The firm sets price and quantity to maximize revenue net of operating cost while ensuring that (i) price does not exceed the price cap, (ii) quantity does not exceed capacity and (iii) quantity does not exceed demand realized at the chosen price:

$$\text{Maximize} \quad (p - v)q \tag{8}$$
$$\text{Subject to: } p \leq \hat{p},\ q \leq \hat{q},\ q \leq Q(p)$$

Of course the firm will shutdown whenever $\hat{p} < v$.

4. FULL COMMITMENT: OPEN-LOOP EQUILIBRIA

Here we adopt an extreme view of commitment: suppose that both the regulator and the firm simultaneously select an entire sequence of actions at the outset, and never deviate from their plans. This type of commitment, though unrealistic, is examined partly to isolate the effects of opportunistic behavior on the outcome.

An open-loop strategy for the regulator is a sequence of price caps chosen given the firm's sequence of investments. Likewise, the firm proposes an investment plan given the sequence of price caps. Except for the irreversibility constraint (i.e., $\hat{q}_t \geq \gamma\hat{q}_{t-1}$), their choices at any date are independent of the history of the game. Starting from any initial capacity for the system, we find a unique Nash equilibrium of this game such that, for each t, $\hat{p}_t = v$ and $\hat{q}_t = \gamma^{t-1}k_0$. Consequently the only steady-state equilibrium capacity is the trivial sequence $\bar{q} = 0$.

First we verify that this is an equilibrium. Note that if $\hat{p}_t = v$ for each t then the firm would never wish to invest because it will not earn any return on capital. Obversely, if $\hat{q}_t = \gamma^{t-1}k_0$, then the regulator will surely have no incentive to provide the firm with a return on investment. Next we show there is no other equilibria. Suppose $\hat{p}_t > v$ for some t, say t'. Then if t' > 0, the regulator can increase its period t' payoff, $S(\hat{p}_{t'}, \hat{q}_{t'})$, by reducing $\hat{p}_{\bar{t}}$ down to v. On the other hand, if $\hat{q}_{t'} = 0$ then reducing $\hat{p}_t$, to v has no effect on the regulator's payoff, or more succinctly, $\hat{p}t = v$ is a best reply to $\hat{q}_t \geq 0$ for each t. Thus the only equilibrium has no investment and price caps equal to unit operating costs in each period.

Proposition 1: *For any initial capacity, k_0, there is a unique open-loop equilibrium: $\hat{p}_t = v$ and $\hat{q}_t = \gamma^{t-1}k_0$ for each $t \geq 1$.*

At any open-loop equilibrium the firm will not invest as long as $\delta < 1$. Therefore, if initial capacity is positive it will be allowed to decay along the equilibrium path. Whence, the only steady-state open-loop equilibrium for $\delta < 1$ is the trivial case when initial capacity is zero.

Corollary 2: There is a unique steady-state equilibrium: $\hat{q}_t = 0$ and $\hat{p}_t = v$ for all t.

These conclusions are sensitive to the rule for determining final price and quantity. This is clear if we consider an extreme rule for settling the indeterminacy. Suppose that, whenever installed capacity is insufficient to satisfy all demand realized at the price cap, the firm is forced to shut down: $p(\hat{p},\hat{q}) = \hat{p}$ and $q(\hat{p},\hat{q}) = 0$ if $\hat{p} \geq P(\hat{q})$. Under this "shutdown" rule a wide range of values $(\bar{p},\bar{q})$ where $\bar{p} = P(\bar{q})$ are potential equilibria.

To see this consider $\bar{q} \epsilon [\hat{q}_m, \bar{q}^*(\delta)]$ where $\bar{q}^*(\delta)$ is the optimal capacity given the depreciation rate δ and $\hat{q}_m = \text{argmax}\{[P(q)-v]q - C(\delta q)\}$ is the *monopoly* steady-state capacity. Let $\bar{p} = P(\bar{q})$ and suppose $k_0 = \bar{q}$. Consider the following strategy pair: $\hat{p}_t = \bar{p}$ and $\hat{q}_t = \bar{q}$ for each t = 1,2,... We claim this is an equilibrium as long as steady-state short-run profit:

$$\bar{\pi}(q) \equiv [P(q) - v]q - C(\delta q) \tag{9}$$

is decreasing on $[\hat{q}^m, \bar{q}^*]$. Producing more than $\bar{q}$ and lowering price is not in the interest of the firm by assumption that $\bar{\pi}'(q) \leq 0$. Profits are nonnegative at $\bar{q}$ and so the firm cannot increase its profits by shutting down (which is the consequence of allowing $\hat{q}_t$ to shrink below $\bar{q}$). Additionally, the regulator cannot hope to increase surplus by raising $\hat{p}_t$ at any t, and further reductions in $\hat{p}_t$ at any date t will result in $q_t = 0$ so the regulator will not try to reduce the price cap.

Consequently virtually any steady-state capacity and market-clearing price is a potential equilibrium under the shutdown rule. It confers upon the firm a strong bargaining position by pressuring the regulator to avoid service interruptions.

Proposition 3: *If the firm must shut down when capacity is insufficient, then any strategy pair $\hat{p}_t = \bar{p}$ and $\hat{q}_t = \bar{q}$, for $\bar{q} \in [\hat{q}^m, \hat{q}^*]$, for each $t = 1,2,\ldots$ is an open loop equilibrium as long as $k_0 = \bar{q}$, and provided $\bar{\pi}(q)$ is decreasing on $[\hat{q}^m, \hat{q}^*]$.*

5. NO COMMITMENT: TRIGGER PRICE EQUILIBRIA

In this Section we drop all pretense that the firm or the regulator is able to commit to a future course of action. This is captured by the use of strategies that are contingent on the history of play, i.e., closed-loop strategies. We show that under the delegation rule, the best steady-state equilibrium from the regulator's perspective depends on the discount and depreciation rates, as long as the choices of $\hat{q}_t$ and $\hat{p}_t$ are made simultaneously. This implies that the steady-state planning solution is not an equilibrium. However, as the discount and depreciation rates tend to zero, the best supportable steady state converges to the planning solution. Further, when the firm's choice of $\hat{q}_t$ precedes the regulator's choice of $\hat{p}_t$ in each period, then the planning solution is an equilibrium.

5.1 Best Attainable Steady State Equilibria with Simultaneous Moves

The best achievable steady-state capacity as a closed-loop equilibrium outcome depends on the interest rate, r, and the depreciation rate δ, and is denoted $\hat{q}(\delta,r)$. Define $\hat{q}(\delta,r)$ as the largest $\hat{q}$ such that:

$$\bar{\pi}(\hat{q}) = [P(\hat{q}) - v]\hat{q} - C(\delta\hat{q}) \ge \gamma r C[\delta\hat{q}]/(1 - \gamma r) \qquad (10)$$

and for each $\tau \ge 1$:

$$\sum_{t=\tau}^{\infty} \rho^t S[P(\hat{q}),\hat{q}] \ge \sum_{t=\tau}^{\infty} \rho^t S(v,\gamma^{t-\tau}\hat{q}) \qquad (11)$$

Note that given $\delta < 1$ both inequalities (10) and (11) are satisfied for ρ sufficiently close to one. Thus, as $\rho \to 1$, $[P(\hat{q}) - v]\hat{q} - C(\delta\hat{q}) \to 0$ and so $\hat{q}(\delta,r) \to \bar{q}^*(\delta)$, the optimal steady-state level of capacity. In the following fix δ and let $\bar{q}^* \equiv \bar{q}^*(\delta)$. The derivation of (10) is discussed below.

Given $k_0 = \bar{q}^*$, the outcome path $\hat{q}_t = \bar{q}^*$, $\hat{p}_t = \bar{p}^*$ for all t *cannot* be a subgame perfect equilibrium outcome. To see this note first that the firm's payoff along the path is 0. (Note that at the steady state optimum, $p_t q_t = \bar{p}^*\bar{q}^* = v + C(\delta\bar{q}^*)$ for each t and so profit is zero in each period.) In contrast, the firm can earn a positive profit by deviating from the optimal path, independent of the regulator's actions following a deviation by the firm. Without loss of generality, suppose the firm defects in period 1. By setting capacity at $\hat{q}_1 = \gamma k_0 = \gamma\bar{q}^*$ the firm earns a profit in period 1 of $\gamma C(\delta\bar{q}^*)$ assuming the regulator sets the price at $\bar{p}^*$. And further, by reducing subsequent investment to zero, and letting its capacity depreciate, the firm can prevent future losses independently of the actions taken by the regulator. Hence, this defection is always better for the firm than conforming to the optimal path. In the case Π^0, the defection paths for the firm and the regulator are illustrated in Figure 2 and the results stated in:

Proposition 4: The optimal planning solution is not a closed-loop equilibrium outcome in the time-dependent supergame between the regulator and the firm.

Before describing the equilibrium strategies that result in the capacity as defined by (10) and (11), some additional notation is needed. The history of the game up to period t is denoted as:

$$h_t = (\hat{q}_0, \hat{q}_1, \hat{p}_1, \hat{q}_2, \hat{p}_2, \ldots, \hat{q}_{t-1}, \hat{p}_{t-1}) \tag{12}$$

A strategy for the regulator specifies a price cap in each period t as a function of the history up to that time:

$$\phi_t : h_t \rightarrow \hat{p}_t \in R^1_+ \tag{13}$$

Similarly, a strategy for the firm specifies a capacity for each period t as a function of the history:

$$\psi_t : h_t \rightarrow \hat{q}_t \in R^1_+ \tag{14}$$

The sequences $\phi = \{\phi_t : t=1,2,\ldots\}$ and $\psi = \{\psi_t : t=1,2,\ldots\}$ constitute strategies for the closed-loop game. Consider the following closed-loop strategies which have a "trigger":

$$\phi^*_t(h_t) = \begin{cases} p(\delta,r) & \text{if } h_t = h^*_t \\ v & \text{otherwise} \end{cases} \tag{15}$$

$$\psi^*_t(h_t) = \begin{cases} \hat{q}(\delta,r) & \text{if } h_t = h^*_t \\ \gamma q_{t-1} & \text{otherwise} \end{cases} \tag{16}$$

where $\hat{q}(\delta,r)$ is as defined by (10) and (11), $\hat{p}(\delta,r) = P[\hat{q}(\delta,r)]$ is the corresponding price, and h^*_t is the sequence $\hat{p}_\tau = p(\delta,r)$ and $q_\tau = \hat{q}(\delta,r)$ for each $t = 1,2,\ldots,t-1$. Then the strategy pair (15), (16) is a subgame perfect Nash equilibrium if at any point in the game and for any history up to that point, both the regulator and the firm would wish to continue using those strategies (assuming the other one will do so as well).

Faced with a defection at any time on or before period $t-1$, the regulator will set $\hat{p}_t = v$, as it anticipates that the firm will make no further investment independent of its own decision. Secondly, given no previous defection up through date τ, (2), (15) and (16) implies that defection by the regulator does not pay provided:

$$\sum_{t=\tau}^{\infty} \rho^t S[\hat{p}(\delta,r),\hat{q}(\delta,r)] \ge \sum_{t=\tau}^{\infty} \rho^t S[v,\gamma^{t-\tau}\hat{q}(\delta,r)] \tag{17}$$

But, this is guaranteed by (11).

Now, once a deviation from h_t^* has occurred, the firm will never wish to incur additional investment expenditures as it anticipates the regulator setting $\hat{p}_t = v$ for all subsequent periods. Given $h_t = h_t^*$, the maximum one-time profit the firm can obtain is:

$$(\hat{p}_t - v)\gamma\,\hat{q}(\delta,r) = \gamma C[\delta\hat{q}(\delta,r)]/(1 - \gamma r) \tag{18}$$

where, for ρ sufficiently close to one:

$$\hat{p}_t = v + C[\delta\hat{q}(\delta,r)]/[(1 - \gamma r)\hat{q}(\delta,r)] \tag{19}$$

as in (10). And for such ρ, the profits from continued cooperation is:

$$\sum_{t=1}^{\infty} \rho^t \Big\{[p(\delta,r) - v]\hat{q}(\delta,r) - C[\delta\hat{q}(\delta,r)]\Big\} \tag{20}$$

For ρ sufficiently near 1, (10) implies that this can be rewritten as:

$$\{C[\delta\hat{q}(\delta,r)]/(1 - \gamma r) - C[\delta\hat{q}(\delta,r)]\}/r = \gamma C[\delta\hat{q}(\delta,r)]/(1-\gamma r) \tag{21}$$

and this last expression is the same as (18). Therefore, the firm cannot gain from defection when ρ is sufficiently large. Note, too that for larger values of $\hat{q}$, (10) would be violated for the firm (provided ρ is sufficiently close to 1) and thus, no larger steady-state capacity is sustainable as a subgame perfect equilibrium outcome. Thus, the strategies (ψ^*,ϕ^*) result in the welfare-maximal feasible steady-state outcome.

Furthermore, once the trigger has been set off, the punishment path is a Nash equilibrium in the continuation subgame. This is so since the regulator has no incentive to allow further profits for the firm if the firm is undertaking no new investment; the firm has no incentive to incur additional investment expenses if the regulator is going to set low rates.

Finally, it is important to note that our trigger strategies call for minimax punishments of defections. As a result outcomes supported by these strategies cannot be Pareto dominated by any other pair of closed-loop strategies. Summarizing the above arguments gives us:

Proposition 5: Given any $\delta < 1$, $\exists\ \bar{r} \ni$ for $r \le \bar{r}$, the trigger strategies, (ψ^*,ϕ^*), of (15) and (16) are a subgame perfect equilibrium. Furthermore, the equilibrium outcome from play of (ψ^*,ϕ^*) is the best feasible steady-state outcome.

It can be shown that for $\delta = 1$ the trigger strategies fail to support the optimal steady-state outcome. Consider any $q^+ > 0$ such that $p^+ = P(\hat{q}^+) > v$. If the firm's trigger strategy calls for it to set $\hat{q} = q+$ as long as $\hat{p}_s = p^+$ for all $s < t$, then the regulator will not wish to maintain $\hat{p}_t = p^+$. For (p^+, q^+) to be a steady-state equilibrium outcome, (11) must hold. But this cannot be the case when $\delta = 1$ since then capacity does not depreciate. Further, the nearer δ is to 1 the smaller the discount rate must be in order for the regulator to comply with any steady state $\bar{p}^* > v$.

5.2 Equilibrium where the Firm Commits before the Regulator.

Now we suppose that the firm must choose $\hat{q}_t$ before the regulator sets $\hat{p}_t$. In this game a strategy for the firm is, as before, a rule determining choice of capacity in each period t as a function of the history through period (t-1). However, a strategy for the regulator specifies the choice of $\hat{p}_t$ as depending on h_t and $\hat{q}_t$, the firm's capacity in period t. Let $h_t^+ = [h_t, \hat{q}_t]$. And consider the following strategy pair (ψ^*, ϕ^*):

$$\psi_t^*(h_t) = \begin{cases} \hat{q}^* & \text{if } h_t = h_t^* \\ \gamma \hat{q}_{t-1} & \text{otherwise} \end{cases} \tag{22}$$

$$\phi_t^*(h_t^+) = \begin{cases} \hat{p}^* & \text{if } h_t^+ = (h_t^*, \hat{q}_t^*) \\ v & \text{otherwise} \end{cases} \tag{23}$$

where $h_t^* = (\hat{q}^*, \hat{q}^*, \hat{p}^*, \hat{q}^*, \hat{p}^*, \ldots, \hat{p}^*)$ (the optimal path through period t-1). Given $\delta < 1$, this strategy pair is a subgame perfect equilibrium which attains the optimal planning solution provided r is not too large. The regulator will cooperate if the discount rate is not too large. The firm earns zero profit along the equilibrium path, and since it moves first, it cannot earn more by setting off the trigger. But once a deviation has occurred, neither player has an incentive to cooperate.

***Proposition 6:** Given $\delta < 1$, there is a $\bar{\rho} < 1$ such that the optimal planning solution is an equilibrium outcome whenever $\bar{\rho} \le \rho \le 1$.*

6. THE LINEAR-QUADRATIC CASE

A linear-quadratic example not only can illustrate the above steady-state results but it also can extend the results to cover equilibrium along a dynamic path. Let the inverse demand be:

$$P(q) = a - bq \tag{24}$$

so that the price elasticity is $\varepsilon_t = (a - bq_t)/bq_t$. Consumer surplus then becomes:

$$S(p,q) = aq - bq^2/2 - pq \tag{25}$$

which equals $bq^2/2$ whenever $p = a - bq$. Investment cost is given the general quadratic form:

$$C(I) = \alpha + \beta I + \xi I^2/2 \tag{26}$$

We continue to assume constant marginal operating costs of v so that dynamic marginal cost in (6) reduces to:

$$MC_t = v + \beta\rho(\delta + r) + \xi(I_t - \gamma\rho I_{t+1}) \tag{27}$$

All parameters a, b, v, α, β and ξ are assumed positive; we also require sufficient demand to cover total average cost in a steady state, or $(a - bq)q > vq + (\alpha + \beta\delta q + \xi\delta^2 q^2/2)$ for some q. Other restrictions will be indicated as needed.

In principle, we are in a position to compute solutions to the various outcomes given the demand and cost paramenters and the initial values (k_0, Π^0). We are particularly interested in sweeping out the combinations of surplus and profit $(\mathcal{S}, \Pi^0)$ that can be supported by trigger strategies for different values of depreciation and interest rates, i.e., γ and ρ.

Before turning to the various solutions, we describe the firm's short-run behavior. Within each period, the firm maximizes current profit faced with a price cap and a capacity level. In the quadratic case this becomes:

$$(p_t, q_t) = \begin{cases} ((a+v)/2,\ (a-v)/2b) & \hat{p}_t \geq (a+v)/2,\ \hat{q}_t \geq (a-v)/2b \\ (a-b\hat{q}_t, \hat{q}_t) & \hat{p}_t \geq a-b\hat{q}_t, \quad \hat{q}_t \geq (a-v)/2b \\ (a-b\hat{q}_t,\ \hat{q}_t) & a-b\hat{q}_t \leq \hat{p}_t \leq (a+v)/2 \\ (\hat{p}_t,\ \hat{q}_t) & a-b\hat{q}_t \geq \hat{p}_t \geq v \\ (\hat{p}_t,\ 0) & \hat{p}_t < v \end{cases} \tag{28}$$

Thus, the firm engages in monopoly pricing when the price cap and capacity constraint are slack as in the first range. In the second and third range, the firm is capacity constrained and price constrained, respectively. Both constraints bind when there is insufficient capacity to meet realized demand at the price cap. And finally, the firm shuts down when price falls below variable cost -- though it may still maintain a stock of capital.

6.1 The Planning Solution

Substituting the expressions for p_t, ε_t and MC_t into the pricing rule (5) yields a second-order, non-homogeneous, linear difference equation:

$$a - v - \beta(\delta + r)\rho - (1 + \eta)bq_t + \zeta[\gamma\rho\hat{q}_{t+1} - (1 + \gamma^2\rho)\hat{q}_t + \gamma\hat{q}_{t-1}] = 0 \tag{29}$$

When initial capacity is smaller than the steady-state optimum (i.e., $k_0 < \bar{q}^*$), there is no excess capacity (i.e., $q_t = \hat{q}_t$), and so (29) reduces to:

$$A\hat{q}_{t+1} + B\hat{q}_t + C\hat{q}_{t-1} + D = 0 \tag{30}$$

where $A = \zeta\gamma\rho$, $B = -(1 + \eta)b - \zeta(1 + \gamma^2\rho)$, $C = \zeta\gamma$, $D = a - v - (\delta + r)\rho\beta$. A closed-form solution of this difference equation can be derived using standard procedures; this is done in the Appendix. The optimal accumulation path is found to be:

$$\hat{q}_t^* = \lambda^t k_0 + (1 - \lambda^t)\bar{q}^* \tag{31}$$

where $\lambda = \{1 + \eta + \zeta(1+\gamma^2\rho) + \surd(1+\eta)^2 + 2(1+\eta)\zeta(1+\gamma^2\rho) + \zeta^2(1-\gamma^2\rho)^2\}/2\zeta\gamma\rho$ which lies in the unit interval. Thus each member of the capital stock sequence lies between the initial and the steady-state capacities. The sequence is monotone increasing up to the steady state and exhibits the Markov property such that investment in any period depends solely on last period's capital stock.

Setting $\hat{q}_{t+1} = \hat{q}_t = \hat{q}_{t-1}$ in (30) yields the steady-state capital stock:

$$\bar{q}^* = [a - v - (\delta+r)\rho\beta]/[(1+\eta)b + \zeta(1+\gamma^2\rho) - \zeta\gamma(1+\rho)] \qquad (32)$$

which depends on the budget constraint Π^0 through the Ramsey number η. Since $r\Pi^0 = \sum_1^\infty \rho^t[(\bar{p}^* - v)\bar{q}^* - C(\delta\bar{q}^*)]$ in the steady state, $\bar{q}^*$ can be written as a function of Π^0, and in this case solves a quadratic equation.

When initial capacity is "large" (i.e., $k_0 > \bar{q}^*$), the solution calls for no investment until the first time $\bar{t}$ that capacity falls to the optimal steady-state level: $\gamma^t k_0 \le \bar{q}^*$. After that time, the firm invests to maintain the steady-state capacity. While approaching the steady state, price is set at operating cost $p_t = v$, unless there is insufficient capacity (i.e., $\hat{q}_t < Q(v)$), in which case price just clears the market $p_t = P(\hat{q}_t)$. Below we will confine our attention to the case in which $k_0 < \bar{q}^*$. In addition, we will assume α is sufficiently small (relative to the depreciation rate) so that it is optimal to invest in every period.

6.2 Trigger Price Equilibria Along the Dynamic Path

We now construct the closed-loop equilibrium supported by trigger strategies when the initial capital stock differs from that of the optimal steady state. Given the efficient path of capacities, $\hat{q}_t^*$, consider the following trigger strategies:

$$\phi_t = \begin{cases} \hat{p}_t^* & \text{if } \hat{q}_s = \hat{q}_s^*,\ \hat{p}_s = a - b\hat{q}_s^*, \quad s = 1,\dots,t-1 \\ v & \text{otherwise} \end{cases} \qquad (33)$$

$$\psi_t = \begin{cases} \hat{q}_t^* & \text{if } \hat{q}_s = \hat{q}_s^*,\ \hat{p}_s = a - b\hat{q}_s^*, \quad s = 1,\ldots,t-1 \\ \gamma\hat{q}_{t-1} & \text{otherwise} \end{cases} \tag{34}$$

We will compare the relative merits of defection and cooperation first by the firm, and then by the regulator. Without loss of generality, suppose that defection occurs in period $t = 1$, and use a tilde $\tilde{}$ to distinguish variables along the defection path. The defection paths for both the firm and the regulator are inscribed in Figure 3.

Defection by the Firm: The firm enters period 1 with capacity $\gamma\hat{q}_0$ and an unsuspecting regulator sets the price cap at $\hat{p}_1^* = a - b\hat{q}_1^*$. The firm knows that price will be cut to operating cost $\tilde{p}_t = v$ once it deviates, at which time its best reply is to cease all investment; thereafter, $\tilde{q}_t = \gamma^{t-1}\tilde{q}_1$ where $\tilde{q}_1 = \gamma k_0 + \tilde{I}_1$. Initial capacity could be so small that it nevertheless pays to make some investment $\tilde{I}_1$ in the defection period. Generally, the firm will not find it in its best interest to do so. In that case the maximum profit the firm can obtain in period 1 is $(\hat{p}_1^* - v)\gamma k_0$, followed by no operating profit in subsequent periods. Since the firm receives profit π^0 at the planning solution, the net gain of defection over cooperation is:

$$\begin{aligned} \Delta\pi &= \tilde{\pi} - \pi^* \\ &= (\hat{p}_1^* - v)\gamma k_0 - \pi^0/\rho \\ &= \{a - v - b[\lambda k_0 + (1 - \lambda)\bar{q}^*]\}\gamma k_0 - \pi^0/\rho \\ &= Ak_0^2 + Bk_0 + C \end{aligned} \tag{35}$$

where $A = -b\gamma\lambda$, $B = a - v - b(1 - \lambda)\bar{q}^*$, $C = -\pi^0/\rho$. Thus, given π^0, defection is profitable (i.e., $\Delta\pi > 0$) when:

$$k_0 \in \left(\frac{-B - \sqrt{B^2 - 4AC}}{2A}, \frac{-B + \sqrt{B^2 - 4AC}}{2A}\right) \tag{36}$$

This interval vanishes whenever $\Pi^0 \le \rho\big(a - v - b(1 - \lambda)\bar{q}^*\big)^2/4b\gamma$, so that defection by the firm is guaranteed when it is offered an insufficiently small market value at the planning solution.

Defection by the Regulator: If instead the regulator considers defection in the first period, cutting price down to operating cost $\tilde{p}_1 = v$ is a best reply strategy because one-period consumer surplus is strictly decreasing in price. The firm will add the efficient amount of capital in the first period $\tilde{q}_1 = \hat{q}_1^*$, but since no capital recovery will be provided after that time, it will stop investing, and hence $\tilde{q}_t = \gamma^{t-1}\hat{q}_1^*$ for $t \ge 2$. The payoff to the regulator is then:

$$\tilde{\mathcal{S}} = \sum_{1}^{\infty} \rho^t\big[(a - v)\gamma^{t-1}\hat{q}_1^* - b\big(\gamma^{t-1}\hat{q}_1^*\big)^2/2\big] \tag{37}$$

We need to check whether this value exceeds the surplus obtained from cooperating, namely:

$$\mathcal{S}^* = \sum_{1}^{\infty} \rho^t\big[b\,\hat{q}_t^{*2}/2\big] \tag{38}$$

Substituting $\hat{q}_t^* = \lambda^t k_0 + (1 - \lambda^t)\bar{q}^*$ in both (38) and (39), we find that:

$$\Delta\mathcal{S} = \tilde{\mathcal{S}} - \mathcal{S}^* = A'\underline{q} + \underline{q}'B\underline{q}, \tag{39}$$

where $\underline{q}' = [k_0, \bar{q}^*]$, A' is a 1×2 vector of strictly-positive linear coefficients and B is a 2×2 matrix of strictly-negative quadratic coefficients. The inequality $\Delta\mathcal{S} > 0$ circumscribes an elipse in the $(k_0, \bar{q}^*)$ plane. Since $\bar{q}^*$ depends on Π^0, we have again found a defection set in terms of (k_0, Π^0). Finally, the intersection of the regions $\Delta\Pi < 0$ and $\Delta\mathcal{S} < 0$ given in (35) and (36) delineate the values of (k_0, Π^0) for which trigger strategy equilibria exist.

6.3 Steady-State Equilibria

When the system starts off in the optimal steady state, conditions that gaurantee existence are considerably less complicated to derive. Starting with the case of the firm, we know immediately that profit from cooperation calculated at $t = 1$ is just $\Pi^* = \Pi^0/\rho$. If instead it decides to defect at $t = 1$, then profit is easily calculated to be:

$$\tilde{\Pi} = \gamma[\alpha + \beta\delta\bar{q}^* + \xi(\delta\bar{q}^*)^2/2] - (1 + r\delta)\Pi^0 \tag{40}$$

Then the gain to defection is:

$$\begin{aligned} \Delta\Pi = \tilde{\Pi} - \Pi^* &= \gamma[\alpha + \beta\delta\bar{q}^* + \xi(\delta\bar{q}^*)^2/2] - (1 + r\delta)\Pi^0 \\ &= A\bar{q}^{*2} + B\bar{q}^* + C \end{aligned} \tag{41}$$

where $A = \gamma\xi\delta^2/2$, $B = \beta\delta\gamma$ and $C = \alpha\gamma - (1 + r\delta)\Pi^0$. Clearly the net gain is decreasing in long-run profit Π^0 as it should because cooperation becomes more attractive to the firm the higher the return it offers.

The relative benefit of cooperation over defection to the regulator is also straightforward. The payoff to cooperation is:

$$\mathcal{S}^* = \sum \rho^t \left[bq^{*2}/2\right] = bq^{*2}/\ 2(1 - \rho) \tag{42}$$

To compute $\tilde{\mathcal{S}}$, substitute $\tilde{q}_t = \gamma^{t-1}\bar{q}^*$ into (2) to get:

$$\tilde{\mathcal{S}} = \frac{(a - v)\bar{q}^*}{\gamma(1 - \rho\gamma)} - \frac{b\bar{q}^{*2}}{2\gamma^2(1 - \rho\gamma^2)} \tag{43}$$

Therefore, we can write $\Delta\mathcal{S} = A\bar{q}^* + B\bar{q}^{*2}$ where $A > 0$ and $B < 0$, and thus, there is a critical level of steady-state capacity above which the regulator will cooperate. This makes sense since the larger the long-run capacity, and hence, production level, the greater the benefits to consumers of maintaining the relationship. Alternatively, recall that $\bar{q}^*$ and Π^0 are inversely related, so that the regulator will be tempted to renege if cooperation requires it to transfer a large amount of consumer surplus to owners of the firm.

6.4 A Simple Case: Constant Unit Investment Cost, No Operating Expense

Now let us further simplify the example by assuming investment costs are proportional (i.e., $\alpha = \xi = 0$) and operating expenses are negligible (i.e., $v = 0$). Provided there is no excess capacity, which is valid if $k_0 \leq \bar{q}^*/\gamma$, Ramsey-Boiteux pricing rule reduces to:

$$p_t[1 + \lambda(1 - 1/\varepsilon_t)] = \beta\lambda(1 - \rho\lambda) \tag{44}$$

Note that the right-hand side of (45) is independent of t. Thus, output and capacity are constant, $q_t = \bar{q}^* = \hat{q}_t$ for each $t = 1,2,\ldots$ as long as the left hand side is monotone decreasing. Price is constant as well: $p_t = \hat{p}_t = \bar{p}^* = a - bq^*$ for each t. The breakeven condition becomes:

$$\Sigma\rho^t R(q_t) = \bar{p}^*\bar{q}^*/r = \sum \rho^t\beta(\hat{q}_t - \gamma\hat{q}_{t-1}) = \beta\{\delta\bar{q}^*/[r(1+r)] + (\bar{q}^*-\gamma k_0)\} \tag{45}$$

Further, $R(q_1) - \beta(\hat{q}_1 - \gamma k_0) = \bar{p}^*\bar{q}^* - \beta(\bar{q}^* - \gamma k_0)$. This term represents first period profits and is negative if:

$$\bar{q}^* \geq \{(1-\delta)/[1 + \delta/(1-r^2)]\}k_0 \tag{46}$$

i.e., for $\hat{q}_0$ not too large. This means that the optimal plan can emerge as an equilibrium with trigger strategies as long as:

$$k_0 \leq [1 + \delta/(1-r^2)]q^*/(1-\delta) \tag{47}$$

As above, the trigger strategy pair of (33) and (34) is a subgame perfect equilibrium. This is so even where the choice of $\hat{p}_t$ and $\hat{q}_t$ are not simultaneous.[6] Note that the firm's incentives to defect are not positive as long as $k_0 \leq q^*\{(1 + \delta/(1-r^2))/(1-\delta)\}$. Also, where the firm moves first it cannot benefit by defecting. Further, the regulator has no incentive to deviate as long the interest rate is not too low. Summarizing:

Proposition 7: Suppose $k_0 \leq q^*\{(1 + \delta/(1-r^2))/(1-\delta)\}$.

(i) The optimal planning solution is a subgame perfect equilibrium outcome with the strategy pair described in (33) and (34) and where the regulator and the firm move simultaneously for r sufficiently small.

(ii) Where the firm must choose capacity in each period before the regulator determines the price cap in each period and the interest rate is not too small, then the planning solution is an equilibrium outcome.

We now calculate the the values of r and δ for which the regulator will not defect by solving (17). The payoff to the regulator from setting $\hat{p}_t = 0$ for each $t = 1,2,\ldots$ is $\sum_1^\infty \rho^t[S(0,\gamma^{t-1}\bar{q}^*)] = \sum_1^\infty \rho^t a\gamma^{t-1}\bar{q}^* - \sum_1^\infty \rho^t b(\gamma^{t-1}\bar{q}^*)^2/2 = a\bar{q}^*/(1+r-\gamma) - b\bar{q}^{*2}/[2(1+r-\gamma^2)]$. Continuing with $\hat{p}_t = \bar{p}^*$ for all t results in a payoff:

$$\sum_1^\infty \rho^t b\bar{q}^{*2}/2 = b\bar{q}^{*2}/2r \tag{48}$$

This latter term exceeds the former as long as:

$$\bar{q}^* \geq \{[2a(1 + r - \gamma^2)r]/[b(1 + 2r - \gamma^2)]\} \tag{49}$$

Note, that this last expression tends to 0 as r tends to 0.

7. BIBLIOGRAPHIC NOTES

A well-developed, but non-strategic literature exists which addresses many of the same questions as we do here. Much of this effort has been devoted to the design of formulas for adjusting prices over time so as to induce a public utility to produce at the efficient level. The seminal contribution that bears directly on utility regulation is the scheme suggested by Vogelsang and Finsinger (1979). One feature all the models in this literature have in common with ours is the flexible approach to regulation. Much authority is delegated to the firm under the presumption that broad control and intensive scrutiny culminate in the "dead hand of regulation" so often associated with rate-of-return rate-making. Streamlined procedures such as these not only reduce direct administrative costs but also economizes on scarce demand and cost information.

With a couple of exceptions, however, these models totally neglect the long-run incentives of investors or customers, choosing instead to concentrate on the steady-state limit of the price path with some attention to the rate and direction along way. The incidence of allocative inefficiency on the efficient path is ignored. In a rare exception, Sappington (1980) demonstrates how wasteful expenditures can result when the firm behaves strategically when operating under the Vogelsang-Finsinger mechanism.

Another closely related literature is the recent research surrounding proposals for replacing traditional rate-base regulation with a system of price ceilings, especially in the case of telecommunications (e.g., Linhart, Radner and Sinden (1983), Littlechild (1983), Haring and Kwerel (1987), Vogelsang (1988)). This reform envisions regulators imposing a cap on price indices of groups of services over an extended period of time. The price caps would be revised periodically to take into account general inflation and productivity improvements. Particularly relevant for our model is the work of Linhart, et.al. (1983) which, though once again non-strategic, was the first to our knowledge to propose a trigger type pricing rule for the purposes of regulating a natural monopoly.

In several aspects our approach reflects the spirit of the price cap proposals, but at the same time it re-directs the analysis in a number of significant ways. To begin with, our model accomodates regulators' desire to facilitate a settlement negotiated face-to-face between owners of the firm and representatives for its customers. Indeed we are able to formally derive the "win-win" property that nearly every version of "social contract" regulation strives to achieve.

A consensus has emerged from this debate in favor of de-coupling price setting from the firm's observed performance in order to apply an exogenous force for efficient operation. Of course the possibility that the original policy is exactly right is unlikely, and indeed all of the proposals include a timetable to periodically re-visit the policy. When this occurs the firm and the regulator no longer stand at "arms' length" from one another; opportunities for intervenors to manipulate the outcome of the re-contracting arise. Our model rejects as naive the assumption that regulators could resist the temptation to adjust price caps in response to excessively large or small returns. Instead, we fear that each correction they undertake may draw them closer to a policy that tracks the outcome under standard rate-base regulation.

Drawing from the lessons of sequential games, we happily find that the lack of commitment displayed by our regulator may place no limit on its power to achieve an optimum. The power of trigger strategies have been brought to bear on a number strategic relationships that suffer from inefficient static equilibria. Our formulation of the firm-regulator relationship has a couple of novel features that sets it apart from other applications of trigger strategy equilibria.

First of all, due to the irreversibility of capital, the sequence of constituent games is time-dependent: each new play of the game depends on the entire sequence of plays through the current capital stock. Benhabib and Radner (1987) also include time dependence due to a renewable resource in their re-examination of the "tragedy of the commons." A second characteristic that distinguishes our setup is the inherent asymmetry of the players. The strategies and the payoffs of the firm and regulator are qualitatively different. In this respect our setup is similar to Fernandez and Rosenthal (1988), who model sovereign-debt renegotiation as a time-dependent supergame played by a borrower nation and a lender bank.

8. EXTENSIONS

A number of extensions of our basic model deserve consideration, the most intriguing being the intermediate case of partial commitment. Taking a cue from the work of Reinganum and Stokey (1985), this could be accomplished by allowing the firm and the regulator to adhere to a price and investment paths for a long but finite span of time. Re-contracting would take place between the commitment periods.[8] The value of commitment could then be measured by varying the length of this period, where a length of one period corresponds to our closed-loop game, and an infinite length to the open-loop game.

While trigger strategies have many attractive properties, the penalties they impose are unrealistically severe: firms rarely halt all investment and regulators usually provide some return on invested capital. Less severe penalties can be devised which remain credible and which continue to sustain cooperative outcomes. One approach worth investigating is how the structure of the equilibrium is affected by the length of the penalty period and severity of the penalty. Linhart, et. al. (1983) examined the disciplining effect of lengthening what they referred to as the "recovery period" following a "crisis."

Finally, the specification of the model should be generalized in several directions. Inclusion of multiple products is of interest given that regulated firms typically provide an array of services justified by the scope economies that derive from their capital-intensive technologies. It would also be interesting to allow for investments in cost-reducing innovations. This could be done by permitting operating expense to depend on accumulated expenditures on R&D.

APPENDIX

Derivation of the Solution to the Difference Equation (31):

The standard candidate for a solution to the difference equation (31) takes the form:

$$\hat{q}_t = \phi_+\lambda_+^{\,t} + \phi_-\lambda_-^{\,t} + \bar{q}^* \tag{50}$$

where the λ's solve the quadratic *characteristic equation* $A\lambda^2 + B\lambda + C = 0$:

$$\lambda_{\pm} = [-B \pm \surd\, B^2 - 4AC]/2A \tag{51}$$

It is easy to show that the discriminant is strictly positive so that only real roots exist. From (50) the ϕ's must satisfy the initial conditions $\hat{q}_0 = k_0$ and:

$$\phi_+ + \phi_- = \hat{q}_0 - \bar{q}^* \tag{52}$$

$$\phi_+\lambda_+ + \phi_-\lambda_- = I_1 + \gamma\hat{q}_0 \tag{53}$$

Finally, the expression (32) gives the steady-state capital stock which takes the particular form:

$$\begin{aligned}\bar{q}^* &= -D/(A + B + C) \\ &= [a - v - (\delta+r)\rho\beta]/[(1+\eta)b + \xi(1+\gamma^2\rho) - \xi\gamma(1+\rho)]\end{aligned} \tag{54}$$

As expected $\bar{q}^*$ is decreasing in η, so that as $\eta \uparrow 1$, steady-state capacity drops to the monopoly level.

Boundary conditions are needed to complete the solution. In addition to the initial stock level k_0, the solution converges to the optimal steady state, $\bar{q}^*$. Consequently, both of the first two terms of (50) must vanish in the limit.[9] A simple but tedious calculation shows that $\lambda_+ > 1$ so that necessarily $\phi_+ = 0$, leaving only the last two terms of (50). It can also be shown that $0 < \lambda_- < 1$, and that $\phi_- = \hat{q}_0 - \bar{q}^* < 0$. Suppressing the subscripts on λ_- and ϕ_-, the optimal accumulation path becomes:

$$\begin{aligned}q_t^* &= \gamma^t\hat{q}_0 + (\hat{q}_0 - \bar{q}^*)(\lambda - \gamma^t) + (1 - \gamma^t)\bar{q}^* \\ &= \lambda^t\hat{q}_0 + (1 - \lambda^t)\bar{q}^*\end{aligned} \tag{55}$$

where:

$$\lambda = \{1 + \eta + \xi(1+\gamma^2\rho) + \sqrt{(1+\eta)^2 + 2(1+\eta)\xi(1+\gamma^2\rho) + \xi^2(1-\gamma^2\rho)^2}\} / 2\xi\gamma\rho \qquad (56)$$

A tedious calculation establishes the $\lambda \in (0,1)$.

ENDNOTES

1 Opportunism also arises when, despite promises to the contrary, a regulator succumbs to the temptation to incorporate the firm's private information about technology and demand revealed in the course of its operation. For details, see Baron (1988).

2 This phenomena and its implications are spelled out in detail in Woroch (1986) for the two-period case where construction economies give rise to concave adjustment costs.

3 Since variable cost is independent of capacity, the model cannot easily accomodate cost-reducing innovation.

4 The equality of demand and production is a crucial assumption. By setting unit price p_t different from $P(q_t)$, the planner would be able to make lump-sum transfers in the amount of $[p_t - P(q_t)]q_t$ from investors to customers, thereby solving the firm's budget-balancing problem.

5 A fully general treatment would have price and quantity in each period as a element of the players' strategy sets. No doubt this modification would alter the results.

6 Where $\hat{p}_t$ is set *after* the firm has committed to $\hat{q}_t$ then the planning solution will always emerge as a subgame perfect equilibrium outcome with linear investment costs.

7 In defense of our formulation, we could imagine that the firm threatens to phase out just one of the several services that it supplies. Provided this service is technologically separable from the other, our model applies directly to this more realistic case. Of course if other services are also regulated, the agency may concentrate on its policies in these markets to discipline the firm.

8 Reinganum and Stokey's (1985) take this approach in a model of oligopoly extraction of an exhasutible resource but it is clearly applicable in more general settings. However, by limiting their model to a finite horizon, trigger-type strategies loose much of their appeal.

9 Alternatively, we could choose an initial level of investment I_1 which, along with initial stock $\hat{q}_0$, solved the planning problem. While this approach would produce equivalent results, it is far more cumbersome than pinning down the optimal path with the steady-state condition.

REFERENCES

D. Baron (1988), "Information, incentives, and commitment in regulatory mechanism: regulatory innovation in telecommunications," Stanford University.

W. Baumol, J. Panzar and R. Willig (1982), **Contestable Markets and the Theory of Industrial Structure**, New York: Harcourt, Brace, Jovanovich.

J. Benhabib and R. Radner (1987), "The joint exploitation of a productive asset: a game-theoretic approach," New York University.

W. Brock and W. Dechert (1985), "Dynamic Ramsey pricing," **International Economic Review**.

R. Fernandez and R. Rosenthal (1988), "Sovereign-debt renegotiations: a strategic analysis," NBER Working Paper 2597, May.

J. Friedman (1988), "A modified Folk Theorem for time-dependent supergames," University of North Carolina, February.

V. Goldberg (1976), "Regulation and administered contracts," **Bell Journal of Economics**, Autumn.

J. Haring and E. Kwerel (1987), "Competition policy in a post-equal access market," Federal Communications Commission, Office of Plans & Policies Working Paper 22, February.

P. Linhart, R. Radner and F. Sinden (1983), "A sequential Principal-Agent approach to regulation," Bell Laboratories Working Paper 264, January.

S. Littlechild (1983), "Regulation of British Telecommunications' Profitability," U.K. Department of Industry, February.

J. Reinganum and N. Stokey (1985), "Oligopoly extraction of a natural resource: the implications of the period of commitment in dynamic games," **International Economic Review**, February.

D. Sappington (1980), "Strategic firm behavior under a dynamic regulatory adjustment process," **Bell Journal of Economics**, Spring.

I. Vogelsang (1988), "Price cap regulation of telecommunications services: a long-run approach," Rand Note N-2704-MF, February.

I. Vogelsang and J. Finsinger (1979), "A regulatory adjustment process for optimal pricing by multiproduct firms," **Bell Journal of Economics**, Spring.

O. Williamson (1975), **Markets and Hierachies**.

G. Woroch (1986), "Long-term regulatory contracts," GTE Laboratories Economics Discussion Paper 86-12, February.

FIGURE 1: Extensive Form of the Game

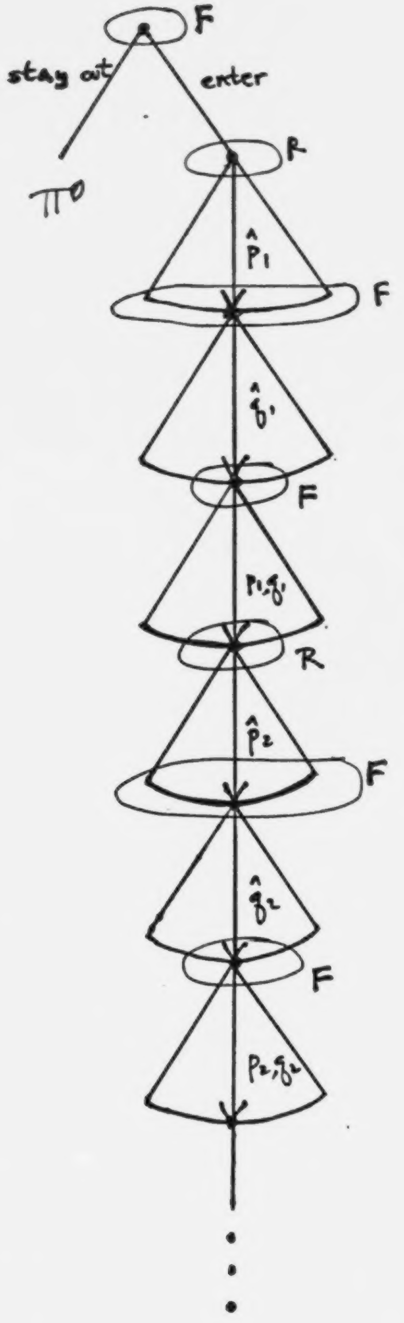

F
stay out
enter
π^0
R
$\hat{p}_1$
F
$\hat{q}_1$
F
p_1, q_1
R
$\hat{p}_2$
F
$\hat{q}_2$
F
p_2, q_2

FIGURE 2: Defection from Optimal Steady-State

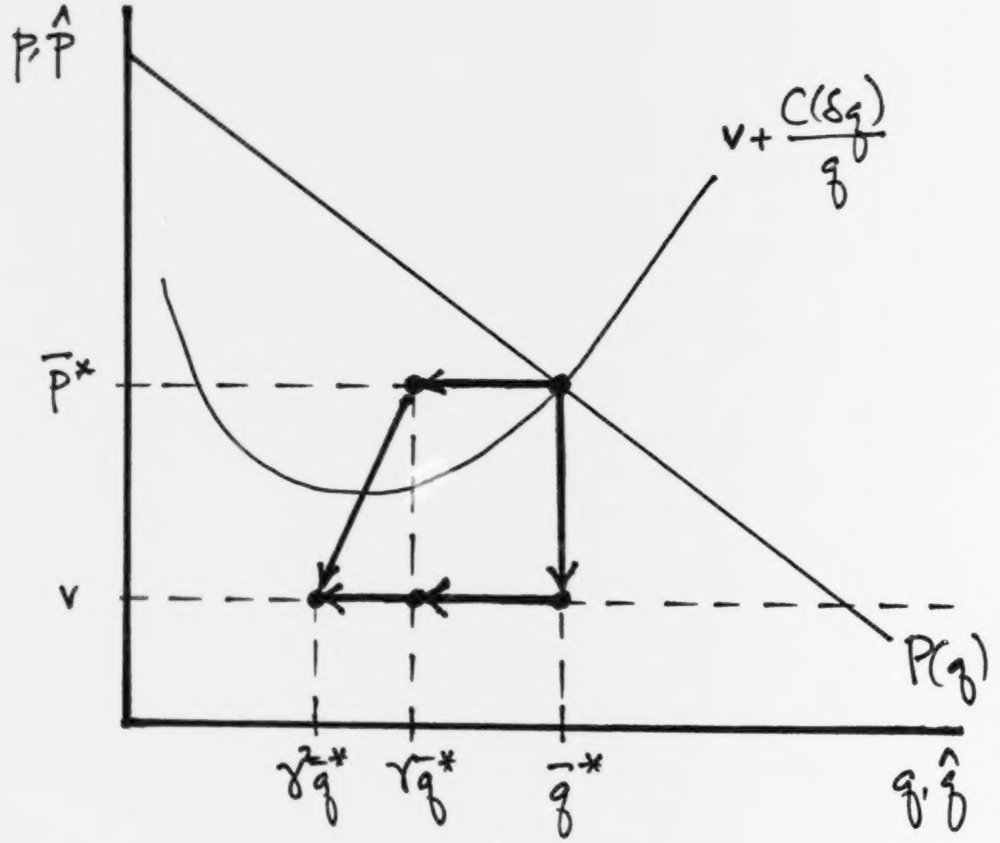
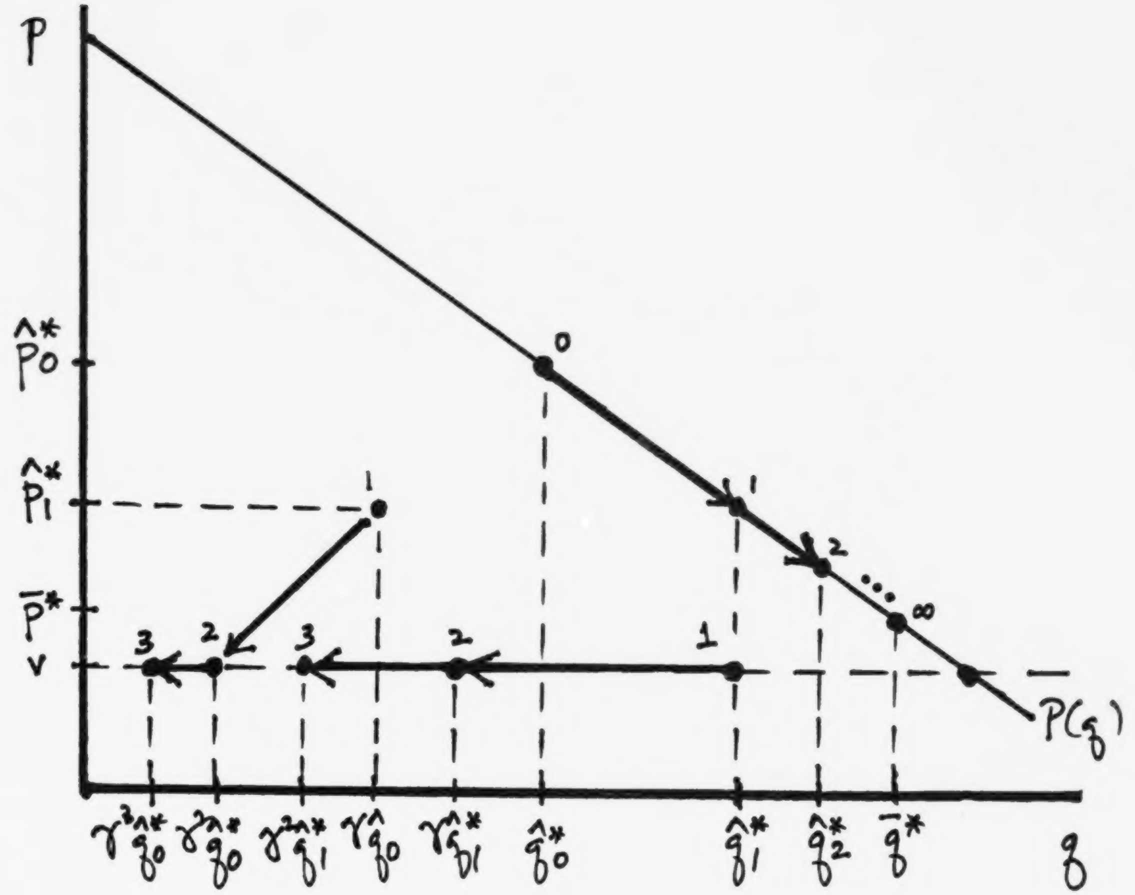

$p, \hat{p}$
$v + \frac{C(\delta q)}{q}$
$\bar{p}^*$
v
$P(q)$
$\gamma^2 \bar{q}^*$
$\gamma \bar{q}^*$
$\bar{q}^*$
$q, \hat{q}$

FIGURE 3: Trigger Price Equilibrium Along the Dynamic Path


P
$\hat{p}^*_0$
$\hat{p}^*_1$
$\bar{p}^*$
v
0
1
2
...
∞
1
3
2
3
2
1
P(q)
$\gamma^3\hat{q}^*_0$
$\gamma^2\hat{q}^*_0$
$\gamma^2\hat{q}^*_1$
$\gamma\hat{q}_0$
$\gamma\hat{q}^*_1$
$\hat{q}^*_0$
$\hat{q}^*_1$
$\hat{q}^*_2$
$\bar{q}^*$
q

BEHIND THE SCENES POWER IN JAPAN'S INFORMATION INDUSTRY

Dr. Jerry L. Salvaggio
Senior Vice President of Information Systems
Whataburger, Inc.
Corpus Christi, Texas

Presented at:
THE SIXTEENTH ANNUAL TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE
Airlie House
Airlie, VA
October 29-30, 1988

ABSTRACT

In this paper, three Japanese organizations, the keiretsu, the sogo shosha and Keidanren are discussed relative to their role in forming information policy in Japan. It will be argued here that while the Ministry of Posts and Telecommunications (MPT) and the Ministry of International Trade and Industry (MITI) are the official policy makers in Japan, the above three organizations have emerged as major agenda setters of informed technology.

BEHIND THE SCENES POWER IN JAPANESE INFORMATION POLICYMAKING

Japanese scholars have argued that American scholars too overemphasize the role the Japanese government plays in developing information technology.[1] The Japanese are no doubt correct. Literature on Japanese technology has focused primarily on the role of the various ministries in guiding the development of information technology. One reason for this is that most information available on Japanese technology is published by the government.

In this paper three Japanese organizations representing private industry, the keiretsu, the sogo shosha and Keidanren are discussed relative to their role in forming information policy in Japan. It will be argued here that while the Ministry of Posts and Telecommunications (MPT) and the Ministry of International Trade and Industry (MITI) are the official policy makers in Japan, private industry has emerged as a major agenda setter of information technology.

---

[1] See Ito, Yoichi, "Recent Trends in the Japanese Telecommunications Policy", in Wedemeyer, D. J. (ed), PTC '83 Pacific Telecommunications Conference: Papers and Proceedings, Honolulu; Pacific Telecommunications Council, 1983.

The above organizations are little known in the U.S. and somewhat opaque to most Japanese. Even if these organizations wished to be better understood they would probably be no less indecipherable. Their modus operandi involves closed meetings, political contributions, bank loans, lobbying, joint ventures and investments. The transparency exists as a result of the way business is conducted in Japan. Tacit agreements are more binding on the players than legal documents. Hierarchy, especially in the keiretsu, is mostly unofficial and personal ties based on common venerable names and university affiliations are often more important than organizational charts.

The first, the keiretsu, operates by virtue of being a large industrial combine involving banking, insurance, shipbuilding, textiles, steel, agriculture and now information technology. The second, the sogo shosha, are trading firms that exercise considerable power through international trade and accumulated wealth. The third, Keidanren, is an economic association whose members come from the keiretsu, sogo shosha and private independent firms. In a sense, Keidanren is the tool or spokesman of the keiretsu and the sogo shosha. All three are influential in the development and adoption of information technology and are capable of setting policy agendas in the information arena.

Keiretsu

Japanese firms realize that to survive within Japan's incredibly competitive environment, and, in foreign markets, they must rely on a strong support system. For some, support comes through membership in one of Japan's powerful industrial groups. Prior to World War II, Japan's hugh financial combines were known as the zaibatsu.[2] The term zaibatsu comes from "zai" meaning money and "batsu" meaning clique.

Most of the original zaibatsu had their origins between the 17th and 19th centuries and were often founded by the samurai or merchant classes. Traditionally, a family held holding company had control over the other members of the zaibatsu and the power that went with it often passed from one generation of the family to the next. Zaibatsu were held together not so much as a result of shared ownership as through central control over banking, transportation and trading. Loyalty to the family in early times, and later to the group itself, formed a bond among members of the group unknown in Western business.

---

[2] An excellent overview of the zaibatsu can be found in Tasuku Noguchi, "The Formation of the Japanese Zaibatsu," _Keio Business Review_, 16, 1979.

Between 1900 and 1945 the zaibatsu were at their zenith in terms of power, influence and accumulated wealth. They were in fact so powerful they were criticized as being a government in and of themselves and a reflection of Western influence. Lucrative government contracts following World War I helped the zaibatsu solidify their powerbase and financial empires. Much of their efforts in the 1930s were turned to developing Japan's industrial capacity for a growing military machine. The military compensated the zaibatsu with prosperous contracts throughout Asia.

When Douglas MacArthur was creating a new architectural design for Japanese society, he immediately targeted the zaibatsu for extinction, believing that they were responsible for fueling the war effort. President Truman's "Basic Initial Post-Surrender Directions to the Supreme Commander for the Allied Powers for the Occupation and Control of Japan" ordered MacArthur to dissolve large industrial and banking combinations. Members of the zaibatsu saw the handwriting on the wall and offered the "Yasuda Plan" in 1945. The plan basically called for the large holding companies to be dissolved.

In time the Yasuda Plan evolved into the "Deconcentration Law", by which MacArthur was able to end the formal structure of the zaibatsu. Leadership of the principle groups was disposed and the stocks of dominate groups sold to the general public on the open market. The Mitsui and Mitsubishi zaibatsu were broken into more than 200 individual firms and investments in Korea, Manchuria and Taiwan were lost.

Had MacArthur understood the deep-felt need of Japanese firms to function in groups, he might have foreseen their eventual reemergence. By the 1970s, the zaibatsu had resurfaced though in a less formal structure which scholars now refer to as "keiretsu." Members of a keiretsu are only loosely tied together since the Deconcentration Law prohibits corporations within a keiretsu from being centrally owned and managed.

Today, the principal keiretsu are Mitsubishi, Mitsui, Sumitomo, Fuyo, Daiichi Kangyo Bank and Sanwa Bank. Each consists of 10 to 30 large corporations, which, in turn, own 50 to 100 percent shares in hundreds of other firms.

One of the more extensive keiretsu in Japan today is Mitsubishi whose origin can be traced back to 1870 when Iwasaki Yataro, a minor samurai, began a special trading firm using sea transportation. In 1888, it expanded into shipbuilding and later established a mining company. Banking was established in 1895. Tokyo

Marine Insurance and Meiji Life Insurance came later.

The Mitsubishi Group enters the 1990s as a closely associated group of immense conglomerates doing business in dozens of unrelated industries. Its member banks are among the world's largest. In telecommunications, it is directly represented by Mitsubishi Electric and indirectly by the Mitsubishi Corporation.

Another powerful keiretsu in Japan today is the Mitsui Group. Its origins can be traced back to the 1600s, further than any family-owned business in the world.[3] Toward the end of the Tokugawa period, the Mitsui family relinquished its ties with the Tokugawas.[4] As a reward for joining the Meiji opposition forces, the group was given directorship of the governmental exchange bureau accounts. In the early 1870s, the family operation expanded and Mitsui joined with the Ono Group to create the Daiichi Wholesale Bank. Four years later, it was changed to the Mitsui Bank with government being the primary depositor.

Mitsui was disbanded after World War II, however, the group bonds were strong and many of the members remained in close contact. In 1950, 27 of the original zaibatsu members re-grouped. As most American businessmen are painfully aware, the new Mitsui Group is quite formidable. Today the group includes 69 firms with approximately 400,000 employees. When the total transactions of all Mitsui Group members are combined, it comes to close to $200 billion.[5] The range of businesses in the group include banking, insurance, trading, mining, real-estate, shipbuilding, and telecommunications. The more prominent members of the group, Toyota, Mitsui Mining, Japan Steel Works, Mitsui O.S.K. Lines and Ishikawajima-Harima Heavy Industries, are among the world's largest corporations. In the area of information technology, Mitsui's major group member is Toshiba, though all the above are now involved in information technology ventures.

---

[3] Most historians trace the origin 1615-1623.

[4] The Tokugawa family ruled Japan from Tokyo between 1600 and 1868 when they were expelled by the Meiji.

Another keiretsu with ancient origins is the Sumitomo Group, formed in the 17th century when Masatomo Sumitomo (1585-1652), a member of the Nehan Sect of Buddhist priests, needed to earn a living to support his family.[6] From a medicine and book shop, the business was expanded considerably several generations later when Masatomo's daughter married the son of Iiemon Soga. Soga had developed a new process for smelting copper which turned out to be revolutionary and important to the development of mining in Japan.

---

[5] The Mitsui Group: Expanding Our Horizons, Internal Publication, Mitsui Public Relations Committee, Tokyo, no date.

[6] Internal publication, Sumitomo publication, N.D.

In the late 1800s, Sumitomo expanded its operations in the West when it purchased a steamship from Great Britain. The family enterprise moved into coal mining and forestry in the early 1900s. Sumitomo Bank was formed in 1895 to provide additional capital for expansion. In 1902, it expanded into steel and in 1934, when heavy machinery was required, the family started Sumitomo Machinery Works.

With the conclusion of World War II, Sumitomo's last Director General, Shunnosuke Furuta, watched the break-up of the Sumitomo empire. Having survived more than three centuries of Japanese history Sumitomo was not to be vanquished so easily. Members of the group quickly formed the Sumitomo Corp., a general trading firm. Today, this firm occupies the center position of the new Sumitomo empire -- an empire which includes the Bank of California.

The Sumitomo Group is involved in information technology in a number of ways. At least seven major electrical and electronic firms, including NEC, are part of the group.

Nature of the Keiretsu

It would be wrong to assume that keiretsu are monopolies. While particular firms within a keiretsu might be part of an oligopoly the group itself cannot be accused of representing an oligopoly or a monopoly as they are made up of many different industries, including banking, trading, electronics, transportation, mining and agriculture.

A keiretsu's power and influence, however, might be instrumental in the emergence of one of the members becoming dominant within a particular industry. NEC, a key firm in the Sumitomo keiretsu is the largest telecommunications firm in Japan but does not hold a monopoly in this area. Whether NEC's success can be attributed to its membership in Sumitomo has not been demonstrated, though, I would argue that NEC's membership in the Sumitomo Group has benefited the firm in immeasurable ways. In 1987, for example, Sumitomo Bank purchased thousands of terminals from NEC.

While the Japanese business world is characterized by large groups not all Japanese firms have a relationship to a keiretsu. Neither Oki nor Sony, for example, are tied to large keiretsu. Hitachi, on the other hand, is loosely tied to DKB, Fuyo and Sanwa Bank.

In theory, members of a keiretsu are not financially related though, in practice, they are intricately tied together. The economic basis of keiretsu activity is more social than financial. The fraternity atmosphere inspires members to work toward common goals. Executives from member firms are able to consort with other executives with the knowledge that they have a common bond.

The full degree to which the keiretsu are involved in the development of Japan's information industry is difficult to ascertain due to the transparency of Japanese business activity. Certain observations however, can be made. First, it is clear that Japan's overall market behavior is influenced by the keiretsu in general. No single keiretsu is capable of changing an industry, or setting policy, but when consensus is arrived among the keiretsu, a new policy agenda can be set that might change market behavior, influence economic trends and serve as a barometer of economic health. The keiretsu are somewhat analogous to bakayashi or Japanese festival music. They set the tone without becoming the center of attention.

Keiretsu facilitate economic activity through control of funds. This is done through the major bank which is at the center of every group. Thus, most of the corporations belonging to a particular keiretsu conduct their commercial transactions and borrow their money from the same banks. Typically, a firm such as Toshiba, which belongs to the Mitsui Group will borrow money from Mitsui Bank and Mitsui Trust as the lead banks and additional funds from banks associated with other keiretsu.

Table 1

MAJOR FIRMS AND ASSOCIATED BANKS AND GROUPS

Firm	Group	Bank
Fujitsu	DKB	Daichi Bank
Hitachi	DKB	Daichi Bank
NEC	Sumitomo	Sumitomo Bank
Mitsubishi Elec	Mitsubishi	Mitsubishi Bank
Toshiba	Mitsui	Mitsui Trust, Mitsui Bank

Member firms of a keiretsu are also able to economize through sharing. Fourteen members of the Mitsui Group, for example, share leased high-speed digital circuits from Nippon Telephone and Telegraph (NTT).

The major technique keiretsu use to the advantage of its numbers is simply to do business with member firms. As one author puts it:

> Mitsubishi companies pay their wages through Mitsubishi Bank, rent their head offices from Mitsubishi Real Estate, and their storage space from Mitsubishi Warehouse. By choice, they use air-conditioning systems made by Mitsubishi Electric, machinery made by Mitsubishi, heavy trucks made by Mitsubishi Motor, and fuel supplied by Mitsubishi Oil.[7]

There are mutually reinforcing connections between two members of the same keiretsu. Employees of members of the same group generally socialize together.[8]

---

[7] Peter Tasker, The Japanese: A Major Exploration of Modern Japan, E.P. Duttan, New York, 1987, p. 261.

Keiretsu are especially adept at coordinating large-scale projects to the advantage of their members. Members frequently cooperate with members of other keiretsu but most often a task is undertaken collectively among members of a single keiretsu. When NTT was privatized, making it possible for new common carriers to emerge, no single firm would have been able to compete as a common carrier due to the capital intensive nature of telecommunication networks. Members of two or three keiretsu, however, were able to pool their resources. Mitsui and Mitsubishi, for example, joined forces to compete against NTT.[9]

The striving for solidarity within a group is facilitated through monthly meetings (kai) of the CEO's of the largest member firms. Usually the meetings take place once a month at a "Nihon Rori" or Japanese traditional restaurant on the same day of the week. This tradition has resulted in each keiretsu referring to its meeting by the day it takes place on. Mitsubishi's monthly meetings are called "Kinyo Kai"

---

[8] Prestowitz tells how employees of Nissan and Hitachi play golf when stationed in London as both are members of the same group. See Prestowitz, Clyde, Jr., Trading Places: How we Allowed Japan to Take the Lead. New York: Basic Books, 1988.

[9] New companies are generally formed by the leading members of a Keiretsu as a Keiretsu itself does not operate as a financial organization.

(Friday meetings) and Mitsui's "Nimo Ku Kai."

The discussions which take place at these meetings are highly confidential with only top executives of the major members being privy to keiretsu strategy. The discussions are also very informal since there is no official hierarchy. Most businessmen in Japan, however, agree that there is an unofficial hierarchy in these monthly meetings that correspond to the annual revenues of the members. Thus, the president of Mitsubishi Electric is likely to have a major voice in the Mitsubishi meetings due to the size and stature of the firm. Generally, the most influential member is a sogo shosha since few firms equal the size of the larger sogo shosha.

Another salient characteristic of keiretsu is that it is not unusual for members of different keiretsu to form partnerships -- especially if the venture is to be highly capital intensive. One of the new common carriers, for example, is backed by Mitsubishi Corp. and Mitsui & Co., the two leading Sogo shosha.

Table 2

JAPAN'S MAJOR KEIRETSU AND ASSOCIATED COMPUTER & TELECOMMUNICATIONS FIRMS

- Mitsubishi
  - Mitsubishi Corp. *
  - Mitsubishi Electric

- Mitsui
  - Mitsui & Co.
  - Toshiba
  - Mitsui Knowledge Industry Co. Ltd.
  - Nishi Nippon Electric Wire & Cable Co.
  - Mitsuigin Software Service Co.
  - Japan Ens

- Fujo
  - Oki
  - Yokogawa Electric
  - Canon
  - Marubeni Electronics
  - Hitachi

- Daiichi Kangyo Bank
  - Furukawa Electric
  - Fuji Electric
  - Fujitsu
  - Hitachi
  - C. Itoh & Co.*
  - Kanematsu Gosho*

- Sanwa Bank
  - Hitachi
  - Sharp
  - Nissho Iwai Co.*

- Sumitomo
  - Sumitomo Electric
  - NEC
  - Sanyo

---

*-sogo shosha

How influential are the keiretsu? I would argue that membership in a keiretsu can be a significant critical success factor in the area of information technology. The reason for this has to do with the capital intensive nature of information technology.

Information technology firms can generally count on low interest loans from member banks and investment partners within the group. This kind of backing provides NEC, Toshiba and others to minimize risk in new capital ventures.

An example of how keiretsu develop new technologies and services was seen in the formation of Japan ENS, one of Japan's new value-added network (VAN) companies. Members of the Mitsui keiretsu lacked experience in information technology.[10] Having decided to enter the VAN market, member companies realized they would need to bring in a firm capable of competing against Fujitsu, NTT and NEC. The logical choice was AT&T. The American firm was eager to enter the market especially in a partnership with the Mitsui Group. Japan ENS was capitalized by AT&T, Mitsui Life Insurance, Mitsui & Co., Toshio Marine and Fire Insurance Co., Dentsu, Nippon Steel Computer Systems, Mitsui Bank, Toyota, Toshiba.[11] With the exception of AT&T, all are members of the Mitsui Group.

---

[10] Toshiba manufactures telecommunications equipment but is weak in voice and data networks.

[11] For a detailed description of Japan ENS, see Jerry L. Salvaggio, the Japanese Computer and Telecommunications Industry: Arlington, VA, Capital Publications, 1986.

Sogo Shosha

Yoshihara Kunio suggests that the term sogo shosha emerged around 1954.[12] The term is used in reference to firms that trade a variety of products, export and import and have offices scattered around the globe. When the keiretsu were dissolved at the end of World War II, the fate of thousands of employees was in question. With the prospect of witnessing the dissolvement of their empire, leaders of the various keiretsu chose to concentrate their resources into one organization. Most chose not a particular industry but the trading firm or sogo shosha. This was logical since the trading firm had always been central to the keiretsu.

Typically, the sogo shosha handle fifty percent of Japan's exports and close to seventy percent of Japan's imports.[13] Large manufacturers, such as Hitachi, also enjoy overseas sales. Sogo shosha, however, are not generally manufacturers.

Today the larger sogo shosha are among the largest firms in the world. Mitsubishi Corp's. total trade transactions make it approximately the same size as IBM.

---

[12] Kunio, Yoshihara, Sogo Shosha, New York: Oxford, 1982.

[13] A detailed analysis of the sogo shosha can be found in Rari Sarathy, "Japanese Trading Companies: Can they be Copied?" Journal of International Business Studies, Summer, 1985, Vo. XVI, No. 2, pp. 101-119.

Mitsui & Co. had total trading transactions of $82 billion in 1986. Everyone of Japan's nine largest trading firms is positioned near the top within their keiretsu. Six sogo shosha are among the ten largest corporations in the world outside of the U.S. All six have greater trade transactions than NTT has revenue.

Sogo shosha can be divided into three categories. In the first category are the "zaibatsu-shosha." These would include Mitsushishi Corp., Mitsui & Co. and Sumitomo. All three are intricately involved in their respective keiretsu.

In the second group are C. Itoh, Marubeni Nisshho Iwai, Tomen and Nichimen. These sogo shosha are associated with various keiretsu but they are not central to a group therefore the Japanese refer to them as "Semi zaibatsu-shosha."

The third category are the "specialized shosha." In this group are Kanematsu-Gosho and several smaller sogo shosha such as Okara Shoji and Toshoku.

Clearly, the most powerful sogo shosha are the "zaibatsu-shosha" such as Mitsubishi, Sumitomo and Mitsui.[14]

---

[14] See Young, Alexander, K. The Sogo Shosha: Japan's Multi-National Trading Companies, Tokyo: Tultle, 1979.

REVENUE OF LARGEST SOGO SHOSHA IN 1988

Table 3

C. Itoh	$72 billion
Mitsui & Co.	82 billion
Mitsubishi Corp.	77 billion
Marubeni Corp.	65 billion
Sumitomo Corp	65 billion
Nissho Iwai Corp.	43 billion

The sogo shosha were at their zenith in the 1960s when Japan was aggressively expanding its exports. When the oil crisis hit Japan the large trading companies suffered dramatically. Some have never fully recovered.

Today, they are clearly in a transition period. Since 1980 they have hired an unusually large number of science and engineering graduates.[15] Realizing that there is little future in shipbuilding, steel, and textiles, the sogo shosha are targeting information technology. As one Japanese writer described the sogo shosha's activities: "Japanese trading companies have recently been absorbed in avoiding becoming dinosaurs by paying very keen attention to developments in the high-tech fields, especially new media.[16]

---

[15] Kiyoshi Otani and Kiyoshi Kusaka, "High Technologies Lure Sogo Shosha as Promising New Fields of Business," The Japanese Economic Journal, April 12, 1982, p. 11.

One of the more aggressive sogo shosha in the Mitsubishi Corp.

Mitsubishi Corp.'s most important step into information technology is clearly its venture into the common carrier market. With Ford Aerospace and Mitsubishi Electric, it launched a satellite in April of 1988 and begin offering services as a common carrier in competition to NTT. Mitsubishi Corp. also markets PBXs in Japan through its venture with Rolm. It has also joined IBM in providing VAN services.

Mitsui & Co.'s glass and steel skyscraper in downtown Tokyo masks its ancestral roots which can be traced back several centuries. It conducts its trading practice on every continent with economic ties into hundreds of areas. Exports to the U.S. alone, comes to more than $12 billion annually, far more than the total exports of many countries. It's world network has centers in Tokyo, Bahrain, London, Sydney and New York and nodes in 88 countries, exchanging 80,000 messages a day.

Like many sogo shosha, Mitsui & Co. has extensive activities in the U.S.[17] Ironically, Mitsui USA, a subsidiary of the parent firm, is actually the U.S.'s fifth largest exporter. A total of $330 million has been invested in American projects either by the U.S. subsidiary or the parent firm in Japan.[18]

---

[16] Kiyoshi Otani, "Sogo Shoshas Shift Away From Goods, Seek Profits In Service Business," The Japan Economic Journal, July 24, 1983, p. 7.

Mitsui & Co. became a major telecommunications player through 50 percent equity in Japan Communication Satellite (JCS) with Mitsubishi Corp. and investing in scores of other information industries. It is also one of a number of partners in NTT International, a high tech organization set up by NTT after privatization. It is also involved in computers through its ownership of Nippon Univac Kaisha Ltd. with Sperry Corp.[19] In the semiconductor area, its subsidiary, Mitsui Bussan Electronics Sales Co., Ltd., produces gallium arsenide, integrated circuits, silicon wafers, and other electronic devices. In 1986, Mitsui & Co., made an equity investment in Shared Technologies, Inc., an American firm with experience in developing intelligent buildings and shared tenant services. More recently the aggressive trading company made a major investment in Tokyo Telemessanger, a paging service in Tokyo.[20]

---

[17] Mitsui & Co.'s trading transactions include machinery, chemical, petroleum, coal, steel, metals, agricultural products,lumber, textiles and general merchandise.

[18] Mitsui USA: The Trading Company, Tokyo: Mitsui & Co., 1986.

[19] Mitsui & Co. owns 34 percent of Nippon Univac.

One of the more aggressive sogo shosha is C. Itoh, founded by Itoh Chubei in 1858, as a traveling wholesale firm. In 1987, it became Japan's largest trading firm moving past Mitsui and Co., and Mitsubishi Corporation, both hurt by the strong yen and falling oil prices. Most of Chubei's early fortunes involved clothing, especially kimono material. When Chubei died in 1903, he had four stores and 100 employees.[21] In 1920, Chubei's son reorganized the family business into an unlimited partnership dealing in both domestic and foreign trade. Though still a major exporter of textiles, Itoh is now scrambling to find its place in telecommunications. C. Itoh is an important member of the Dai-Ichi Kangin industrial group.

In 1985, C. Itoh reorganized its organization and increased its involvement in information technology. It formed a joint venture with Mitsui & Co. and Hughes to purchase and operate Japan Telecommunications Satellite.[22] It operates subsidiaries in the U.S.

---

[20] Mitsui & Co. is not merely an investor but has loaned the firm key management personnel.

[21] Kunio, ibid.

[22] "C. Itoh, Mitsui Joins Hughes to Ford Satellite Business," The Japan Times Weekly, October 6, 1985.

including C. Itoh Electronics, Inc., CIE Terminals, and CIE Systems. C. Itoh uses these firms to export a number of computer peripherals and various electronic products. In 1987, it acquired 33.4 percent of Innotech, a trading company specializing in semiconductor manufacturing equipment.

Possibly, C. Itoh's boldest foray into information technology came with its backing International Telecommunications Japan, an international common carrier that will compete with KDD.

Yet another sogo shosha entering information technology is Marubeni. Marubeni's climb to eminence came slowly. It began as a textile trader in the 1950s, later diversifying into machinery, steel, and telecommunications. It now has more than 170 affiliated companies and equity in over 200 firms, is a major investor in NTT International and has invested heavily in VAN operations. After acquiring 5 percent of the International Mobile Machines Corp. of Philadelphia, Marubeni began marketing the American company's digital telephone equipment. Another subsidiary, Marubeni Electronics, markets computer peripherals.

The Sogo Shosha's New Role

Historically, the sogo shosha's role in the industrialization of Japan was that of importer of raw materials and machinery. Their combined share of imports increased from 56.8 percent in 1979 to 70 percent in 1986. Therefore, it is reasonable to assume that they will continue to be a major source of Japanese imports.

The sogo shosha have not yet made the transition to information technology. Chemicals, fuels, metals, foodstuffs and textiles still account for more than 70 percent of their business. However, to assume that they won't play a key role in the information era would be to ignore their size, experience, resources and recent investment in information technology.[23] When the larger sogo shosha are compared with Japan's largest high tech firms, the latter seem insignificant. The annual revenue for the top six sogo shosha exceed $65 billion each. Fujitsu's annual revenue is less than $20 billion in comparison.

The sogo shosha are increasingly using their enormous cash reserves to fund information technology.[24] In many cases, sogo shosha finance projects using multiple currencies providing the best interest rates. As traders, sogo shosha are very sensitive to economic changes. Product demand is quickly made known through hundreds of branch offices scattered throughout the world.

---

[23] Kunio, ibid.

[24] See for example, Bradley K. Martin, "Japan's Trading Giants Look To Year 2000", The Wall Street Journal, March 31, 1986, p. 22.

Table 4

MAJOR INVESTMENTS OF SOGO SHOSHA

Sogo Shosha	Project	Minor Investor
C. Itoh & Co.	Japan Telecom.	Satellite
	Teleway Japan	
	ITJ	
Mitsubishi	JCS NTT International	Teleway Japan
Mitsui & Co.	JCS	Teleway Japan
	JENS	Nippo Univac
	Tokyo Telemessager	
Nissho Iwai		
Marubeni	International Mobile Machines	
	Marubeni Electronics	

The 1980s will likely be known as a transition era for the sogo shosha. Most continue to lose revenue ranging from 7 percent to 23 percent.[25] They are not able to grow by exporting telecommunication products because private firms have huge marketing arms. They must expand through financial clout and investment in information technology.

Investments since 1985 are staggering. By 1990, the sogo shosha will be using their own satellites, fiber optic networks, and software to conduct financial transactions between Japan and the West and to interconnect dozens of new subsidiaries. They will increasingly use their satellite and fiber optic networks to leverage their businesses in other areas. Sogo shosha are almost certain to form more partnerships with manufacturers.

---

[25] Itoh Grabs Top Trader Rank as Mitsubishi Falls to 5th Place," The Japan Economic Journal, January 31, 1987, p. 3.

Keidanren

In Japan the word "zai" means money and kai means meetings. When the words are combined to form "Zaikai" or money meetings they are in reference to powerful business groups representing different industries. Important zaikai include the Japan Chamber of Commerce and Industry (Nissho), the Japan Federation of Employees Associations (Nikkeiren) and the Japan Committee for Economic Development (Doyukai). All four, however, pale in comparison to Keidanren (Federation of Economic Organizations).[26] Ezra Vogel, in his oft-quoted book, Japan as Number 1, refers to Keidanren as Japan's Foreign Ministry of the Japanese business community.[27] The stature of Keidanren is reflected in the title zaikai sori (prime minister of the business community) given to its president.

---

[26] See Otani Ken, "Enter Keidanren's Young Leaders," Japan Quarterly, 33, October-December, 1986, pp. 368 - 373.

[27] Ezra Vogel, Japan as Number 1

With the end of World War II, Japan's most influential leaders were brought together by the Minister of Commerce and Industry for advise on rehabilitating war-torn Japan. From this group emerged the committee of Economic Organizations, which in 1946, became Keidanren. Its primary goal was to develop consensus on building Japan's economy.

Today, Keidanren is made up of the chief executive officers of Japan's largest industrial companies. It exists for the single purpose of influencing economic policy which it does by constructively channeling money into the Liberal Democratic Party (LDP). Tasker notes that "when top Keidanren executives meet Ministry of Finance officials to discuss economic matters, as they periodically do, the conclusions reached have an import greater than that possessed by LDP policy groups, or sometimes, cabinet decisions."[28]

Within Keidanren members of all the keiretsu and independent firms are intimately linked in common causes. As it turns out, what is good for one industrial group is generally good for another. Thus consensus among the different keiretsu represented in Keidanren is not especially difficult. Advocacy groups within Keidanren study various issues and report back to a central committee where consensus is eventually achieved.

---

[28] Tucker, p. 236.

Keidanren's power comes as a result of its sponsorship of the LDP. Its National Political Association collects monies from its members and forwards it to the LDP. This system has enabled Keidanren to influence elections of prime ministers and economic policy in general. Though it is usually discrete in its use of power, it is known to have played a role in the elections of several prime ministers and in the resignation of at least one prime minister.[29]

Keidanren is often in conflict with the MPT as they have conflicting interests. The MPT seeks to protect existing telecommunications players and users while Keidanren seeks to open markets to new players.

At the vanguard of Keidanren is a very senior group of executives averaging 78 years.[30] As a return for political contributions the party in office creates committees or "shingikai" to study particular problems, often problems identified by Keidanren.[31] Once a shingikai is formed and the problem examined, a report is published.

---

[29] In 1974, Toshiwo Doko, the chairman of Keidanren, called for the resignation of Kakuei Tanaka, the prime minister. Several months later, Tanaka handed in his resignation.

[30] "Old Men Continue to dominate Zaikai World; Osaka-based Businesses Are Unhappy", The Japan Economic Journal, June 26, 1984, p. 46.

[31] Firms belonging to Keidanren are assessed on the basis of their size and profits. The money is then turned over to the Liberal Democratic Party as a political contribution.

Most reports recommend action that is favorable to the business community.

Reports are usually sent to the Diet and/or to the prime minister's office where the chances of an outright rejection are nil as the prime minister was likely to have attended meetings with Keidanren leaders helping to achieve consensus. Once the prime minister is convinced that the action to be taken is in his best interest he forwards it is the appropriate ministry.

In Japan, ministries are autonomous and may or may not accept the recommendations of the prime minister's office. Eventually, most ministries do heed the requests of the prime minister's office -- at least on a theoretical level. Implementation of a major plan, however, is another matter. Ministries are large bureaucracies and, as such, slow to change. Thus, while the minister of the MPT may agree to a request from the prime minister's office, implementation is through thousands of bureaucrats.

Summary

The keiretsu and Keidanren are best viewed as groups which Japanese firms belong to and use for political and financial leverage. Japanese corporations, like Japanese individuals, thrive when operating as a member of a group more than their western counterparts. As Vogel has noted, Americans find groups an imposition while Japanese, from early on, are taught the values of group life.[32] The sogo shosha were highlighted because of their growing prominence in information technology.

"Groups of every other sort abound throughout Japanese society, Reischauer, points out, "and usually play a larger role and offer more of a sense of individual self-identification than do corresponding groups in the United States."[33] Japanese, more than any other people, feel most comfortable when operating within the familiarity of a group.

The group nature of the keiretsu facilitates "consensus," which can then be expressed by Keidanren. Burks offers this observation: "In the Japanese system, it is necessary that leaders (the legitimizer) obtain all the views of subordinates and constituents before announcing a final decision."[34] The goal, he adds, is to retain harmony within the group. I would add that business groups, such as the keiretsu and Keidanren, have secondary objectives, to control the economy, to maximize profits and to influence political and bureaucratic decision making.

---

[32] Vogel, Ezra, Japan As Number 1: Lessons for America, New York: Harper, 1979.

[33] Reischauer, Edwin O., The Japanese, Cambridge: Harvard University Press, 1978, pp. 132-133.

[34] Burks, Ardath, W., Japan: A Pre-industrial Power, Boulder, Westview, 1981, p. 219.

Many important decisions in recent years were initiated by kereitsu members and solidified at keiretsu meetings. These ideas were publicly presented through Keidanren. Keidanren, in turn influenced the LDP. (See Figure 2).

How these three organizations set policy agendas is best illustrated through a brief overview of the privatization of NTT.[35] The initial idea to privatize NTT came from private industry -- particularly the major members of the Mitsubishi and Mitsui keiretsu.

It is highly unlikely that we will ever know who first came up with the idea of privatizing NTT. We can, however, observe certain facts and make certain deductions. We know, for example, that the sogo shosha and particular industries within the keiretsu (steel, textiles, etc.) were experiencing declining profits in the late 1990s and that telecommunications was viewed as the industry of the future. The Japanese were also aware of the possibility that AT&T would be divested opening an entire telecommunications industry to firms long excluded from this arena.

---

[35] See Jerry L. Salvaggio, he Trend Toward Telecommunications Reform: A Comparison of the Divestiture of AT&T and the Privatization of NTT, Keio Communications Review, No. 8, 1987, pp. 55-74.

The idea of privatization evolved into a major issue in the halls of Keidanren.

In the late 1970s there was much excitement in the corridors of Keidanren regarding the potential of information technology. Members had been vigilante in their study of Japan's telecommunications market. Many of the corporations represented in the vaunted organization were beginning to find profits elusive--especially sogo shosha involved in textiles and other sunset industries. Shipbuilding and textile industries were being challenged by South Korea, Singapore, Taiwan and Hong Kong.

Keidanren's Committee on Information and Telecommunication Policy calculated that Japan's telecommunications services market was expanding at an annual rate of 5 percent. This indicated a $20 billion market by 1988. The question was how would Keidanren members enter the market since Fujitsu, Hitachi, Oki and NEC received virtually all NTT contracts for telecommunications equipment.[36]

Rather than compete against the den-den family, most members sought to enter the information industry as investors or sponsors of new companies that could compete against NTT. If newly established common carriers could succeed in capturing 5 percent of the market, they estimated, that would amount to $4 billion in 1988 alone. As it turned out, this is approximately the share the new common carriers achieved by 1988.

---

[36] These firms were, of course, members of Keidanren.

Once the telecommunications market was unlocked in 1985, members of Keidanren capitalized on the new opportunities. Within weeks dozens of new subsidiaries and joint ventures were formed by members of Keidanren. Ironically, many of the new ventures were in the form of partnerships with NTT--the very organization members of Keidanren helped to dislodge from its monopolistic position. In order to create new subsidiaries--NTT needed additional capital, which the more prominent corporate members of the federation could provide. When NTT announced "NTT International" one of its first new ventures, influential members of Keidanren were prominent among the partners. A partial list includes Mitsubishi Corp., Mitsui & Co., Nissho Iwai and Sumitomo.

Members of Keidanren's "Executive Council" were particularly aggressive. Of the ten corporations on the Executive Council, at least seven quickly established joint ventures offering telecommunication carrier services--and three of those six were major backers in Japan's new common carriers.

Figure 1 attempts to show the complexity of Japan's information industry relative to lines of influence. Scholars have tended to focus too much attention and give too much credit to MITI and the MPT. Actually, many of the changes in the information industry and many of the technologies that are developed in Japan are initiated by the keiretsu, promoted by Keidanren and financed by the sogo shosha.

Figure 2 depicts the major forces for and against the privatization of NTT. While the group on the right were entrenched in the telecommunications industry, they were no match for the combined forces on the left.

The major difference between private industry in the U.S. and Japan is that the latter is able to use the keiretsu, Keidanren and the sogo shosha to galvanize leaders in a short period of time. Model 2 shows that the seeds of change and technological development begins with private industry (keiretsu, Keidanren, sogo shosha) and is pushed through by MITI and the MPT. MITI is an especially good catalyst of change.

ALTERNATIVES TO RATE OF RETURN REGULATION IN THE TELECOMMUNICATIONS INDUSTRY

Mark Sievers

US Sprint Communications Company
8140 Ward Parkway
Kansas City, MO 64114
(816) 276-6817

September 30, 1988

Presented at

16th Annual Telecommunications Policy Research Conference
Airlie House, Airlie Virginia

October 30 - November 1, 1988

The materials in this paper are the author's opinions and conclusions. They do not necessarily represent the opinions or policies of US Sprint Communications, its subsidiaries, or other affiliated firms.

ALTERNATIVES TO RATE OF RETURN REGULATION IN THE TELECOMMUNICATIONS INDUSTRY

Mark Sievers

CONTENTS

INTRODUCTION 1

I. **EVALUATION CRITERIA** 5

II. **MINOR REGULATORY CHANGES** 8

- A. **Abbreviated Rate Proceedings** 9
  - 1. Social Contract 10
  - 2. Fast Track Proceedings 11
  - 3. **Analysis** 12
- B. **Rate of Return Bands/Incentive Regulation** 14
  - 1. **Analysis** 15
- C. **Price Bands** 16
  - 1. **Analysis** 17
- D. **Monopoly Services Regulation** 18
  - 1. **Analysis** 21

III. **MAJOR REGULATORY CHANGES** 23

- A. **Revenue Regulation Only** 24
  - 1. **Analysis** 25
- B. **Price Regulation Only** 25
  - 1. Price Caps 26
  - 2. Price Standards 28
  - 3. **Analysis** 30
- C. **Sanctions/Deregulation** 31
  - 1. **Analysis** 33
- D. **Public Ownership** 33
- E. **Franchise Bidding** 34

TABLE 1 EVALUATION CRITERIA

TABLE 2 SUMMARY OF ALTERNATIVES TO RATE OF RETURN REGULATION FOR THE TELECOMMUNICATIONS INDUSTRY

ALTERNATIVES TO RATE OF RETURN REGULATION IN THE TELECOMMUNICATIONS INDUSTRY

Mark Sievers *

INTRODUCTION

Rate of return regulation in the telecommunications industry, as in other public utility industries, refers to the use of revenue requirements to set maximum total revenues and service prices.[1] Generally, revenue requirements are calculated with the following formula:

$$RR = OC + kB$$

where:

RR is the regulated firm's revenue requirements (the maximum revenues that the firm is allowed to earn from regulated operations);

OC is the regulated firm's operating costs, which generally include all prudently incurred current expenses, current depreciation expenses and current taxes;

k is the regulated firm's allowed rate of return; and,

B is the regulated firm's rate base, which is the value of property devoted to public service.

In rate cases, all elements of revenue requirements calculations are open to scrutiny and conflict in adversarial regulatory proceedings. Additional conflict may arise during the rate design portion of regulatory proceedings when individual service prices are set, a process that relies on allocating revenue requirements among services.[2]

---

* Manager — Economic Analysis, US Sprint Communications Company, 8140 Ward Parkway, Kansas City, MO 64114. B.A. (1978 Economics) University of Colorado; M.A., A.B.D. (1982 Economics) University of California, Davis; J.D. (1986) University of Utah. Member Missouri State Bar. Portions of this paper are derived from Sievers, Alternatives to the FCC's Application of Rate of Return Regulation to Local Exchange Carriers, presented at the Rutgers University Advanced Workshop in Regulation and Public Utility Economics, Western Conference, July 6-8, 1988.

[1] There are several excellent texts dealing with rate of return regulation. See generally, A.E. KAHN, THE ECONOMICS OF REGULATION: PRINCIPLES AND INSTITUTIONS (2d 1988) (hereafter cited as KAHN); J.C. BONBRIGHT, A.L. DANIELSON & D.R. KAMERSCHEN, PRINCIPLES OF PUBLIC UTILITY RATES (2d 1988) (hereafter cited as BONBRIGHT); M.A. CREW & P.A. KLEINDORFER, THE ECONOMICS OF PUBLIC UTILITY REGULATION (1986) (hereafter cited as CREW & KLEINDORFER); C.F. PHILLIPS, THE REGULATION OF PUBLIC UTILITIES (2d 1985) (hereafter cited as PHILLIPS); and, L.B. SCHWARTZ, J.J. FLYNN & H. FIRST, FREE ENTERPRISE AND ECONOMIC ORGANIZATION: GOVERNMENT REGULATION (6th 1985) (hereafter cited as SCHWARTZ, et al.). An excellent, concise history of common carrier regulation is in Policy and Rules Concerning Rates for Competitive Common Carrier Services and Facilities Authorizations Therefor: Further Notice of Proposed Rulemaking, CC Docket 79-252, FCC 80-742, Appendix B, DEFINITION OF COMMON CARRIER COMMON LAW BACKGROUND, 84 FCC 2d 445 at 520 (Jan. 16, 1981).

Three distinct rationales or perceptions of revenue requirements are found in the legal and economic regulatory literature. First, revenue requirements are the minimum revenues that a firm should earn to avoid running afoul of Constitutional taking provisions -- revenue requirements are the "just compensation" that a firm must receive for devoting its property to public service.[3] Second, revenue requirements are a means of preventing abuse of monopoly power and are the maximum revenues that a firm should be allowed to earn. Under this perception, revenue requirements are an estimate of the total economic costs of a hypothetical competitive firm, where OC represents the firm's out-of-pocket operating or accounting costs and kB represents an estimate of the firm's opportunity costs.[4] Third, revenue requirements are simply an estimate of the total costs that the regulated firm actually incurs in the provision of service. Under this perception, kB is an estimate of the cost of capital resources used to provide service.[5]

Among commentators, there is a virtually universal disenchantment with rate of return regulation, especially as applied to telecommunications firms. Literally hundreds of legal and economics articles have appeared over the past 20 years that are critical of various aspects of traditional regulation.[6] By comparison, almost nothing has been written in defense of the *status quo*.[7] Much of the criticism of rate of return regulation focuses on its inability to cope with the problems of regulating firms that function in increasingly competitive environments. Other criticisms focus on the administrative inefficiency[8] and distortions of economic incentives[9] created by traditional rate of return regulation. Also, the search for alternatives is greatly influenced by economic dogma -- the belief that regulation is inferior to competition (less regulation) in achieving public policy goals.

---

[2] Broadly speaking, rate of return regulation is only one aspect of utility regulation. The other major aspects include entry regulation, service quality regulation, and exit regulation. Entry regulation is concerned with market structure issues — how many firms should provide service — and is the central issue in certification proceedings. Service quality regulation is concerned with the quality of service provided and usually revolves around service standards. Exit regulation is concerned with the rules regarding discontinuation of services provided to individuals or to entire communities and a common carrier's obligation to serve the public generally.

[3] This rationale is emphasized by courts reviewing regulatory orders. There are a host of decisions in this area. The most well-know include: Bluefield Water Works & Improvement Co. v. Public Service Commission of West Virginia, 262 U.S. 679 (1923) [emphasizing Constitutional restraints on rate making]; Federal Power Commission v. Hope Natural Gas, 320 U.S. 591 (1944) [emphasizing the statutory "just and reasonable" standard]; and, Permian Basin Area Rate Cases, 390 U.S. 747 (1968), reh. den., Bass v. Federal Power Commission, 392 U.S. 917 [emphasizing a "zone of reasonableness"].

[4] This rationale is emphasized by the FCC. See, e.g., Policy and Rules Concerning Rates for Competitive Common Carrier Services and Facilities Therefor: Further Notice of Proposed Rulemaking, CC Docket 79-252, 84 FCC 2d 445 at pars. 15-19 & 33a (Jan. 16, 1981).

[5] This perception is emphasized by firms and regulators in general rate cases. It is also the perception adopted by economists who analyze regulated firms in an Averch-Johnson model of rate base regulation. See generally, Averch & Johnson, Behavior of the Firm Under Regulatory Constraint, 53 AM. ECON. REV. 1053 (1962); and, MURPHY & SOYSTER, ECONOMIC BEHAVIOR OF ELECTRIC UTILITIES, 37-50 (1983).

[6] For surveys of the criticisms of rate of return regulation, see BONBRIGHT, 547-578.

[7] Recent exceptions to this include Cooper, Regulatory Reform in Telecommunications: A Solution in Search of a Problem, 4 TELEMATICS 1 (Nov. 1987); Jones, What's Right With Regulation, 117 PUB. UTIL. FORT. 18 (Mar. 6, 1987); and, Hunt, Why Rate Base Regulation, paper presented to National Association of Regulatory Utility Commissioners, Subcommittee on Telecommunications and Cost Allocation (June 20, 1987).

At a more theoretical level, an increasing number of economists and regulatory experts are challenging traditional perceptions of regulation. In many circles, regulation is not thought of as disinterested, benevolent government intervention to correct for market failures, but rather as the product of "market-like" activities (lobbying, litigation, and advocacy in public forums) undertaken by self-interested stakeholders in the regulatory process.[10] This shift in perceptions has created considerable cynicism regarding traditional regulatory institutions -- rate of return regulation works the way it does (or fails to work) because regulated firms and special interest groups lobbied for and received the regulatory treatment they desired.[11] This school of thought has produced a number of theoretical works by prominent economists and lawyers,[12] and empirical papers that suggest that society is worse off as a result of traditional regulation.[13] For the telecommunications industry, a number of papers have recently appeared that conclude or imply that the imposition of traditional rate regulation results in higher rather than lower telecommunications rates.[14]

---

[8] See generally, S.G. BREYER & R.B. STEWART, ADMINISTRATIVE LAW AND REGULATORY POLICY, 130-180 (2d 1985).

[9] Most of the economic literature in this area concerns the Averch-Johnson-Wellisz hypothesis that rate of return regulation creates incentives to overinvest in rate base assets. For a review and critique of this literature as applied to the telecommunications industry, see Albery & Sievers, The Averch-Johnson-Wellisz Model and the Telecommunications Industry, 40 FED. COMM. L. J. 157 (1988).

[10] A number of commentators have argued that regulators become "captured" by the firms they regulate, and that over time, they tend to view the regulated firm's interests more sympathetically. See generally, Eckert, Life Cycle of Regulatory Commissioners, 24 J. L. & ECON. 113 (1981); Green & Nader, Economic Regulation Versus Competition: Uncle Sam the Monopoly Man, 82 YALE L. J. 871 (1973); and, Wilson, The Dead Hand of Regulation, 25 PUBLIC INTEREST 48 (1971).

[11] See, e.g. Teske, Bypassing the Regulators: US West and State Telecommunications Deregulation, paper presented at Rutgers University Advanced Workshop in Public Utility Economics and Regulation, Western Conference (July 6-8, 1988) [argues that US West has actively and successfully pursued a deregulation strategy in the states it serves]; Ferejohn & Shipan, Congressional Influence on Telecommunications Policy, paper presented at Conference on Utility Regulation, Stanford University Center for Economic Policy Research (April 21-23, 1987) [analysis of Congressional influence on policy development in market entry and the divestiture of AT&T, the FCC's access charge decision, and the line of business restrictions]; and, "Political Entrepreneurs" Lead Many States Toward Reform of Telecommunications Regulation, 5 STATE TELEPHONE REGULATORY REPORT 1 (Oct. 22, 1987).

[12] See, e.g., Demsetz, Why Regulate Utilities, 11 J. L. & ECON. 55 (1968); Posner, Natural Monopoly and Its Regulation, 21 STAN. L. REV. 548 (1969); Stigler, The Theory of Economic Regulation, 2 BELL J. ECON. & MGT. 3 (1971); Posner, Theories of Economic Regulation, 5 BELL J. ECON. & MGT. 359 (1974); Peltzer, Toward a More General Theory of Regulation, 19 J. L. & ECON. 211 (1976); and, McChesney, Rent Extraction and Rent Creation in the Economic Theory of Regulation, 12 J. LEGAL STUD. 101 (1987).

[13] See, e.g., Stigler & Friedland, What Can Regulators Regulate: The Case of Electricity, 5 J. L. & ECON. 1 (1962); Jackson, Regulation and Electric Utility Rate Levels, 45 LAND ECON. 372 (1969); Posner, The Behavior of Administrative Agencies, 1 J. LEGAL STUD. 305 (1972); Moore, Has Electricity Regulation Resulted in Higher Prices? An Econometric Evaluation Utilizing a Calibrated Regulatory Input Variable, 13 ECON. INQUIRY 207 (1975); Meyer & Leland, The Effectiveness of Price Regulation, 62 REV. ECON. STAT. 555 (1980); Moore, U.S. Airline Deregulation: Its Effects on Passengers, Capital, and Labor, 29 J. L. & ECON. 1 (1986); Bailey, Price and Productivity Change Following Deregulation: The U.S. Experience, 96 ECON. J. 1 (1986); Moore, Rail and Trucking Deregulation, in REGULATORY REFORM: WHAT REALLY HAPPENED, 14 (Weiss & Klass, eds. 1986); Primeaux and Mann, Regulator Selection Methods and Electricity Prices, 62 LAND ECON. 1 (1986); and, Nelson & Primeaux, An Examination of the Relationship Between Technical Change and Regulatory Effectiveness, 19 APPLIED ECON. 773 (1987).

What alternatives to traditional rate of return regulation are available? The answer to this question is an increasingly important current policy issue in the telecommunications industry. In 1987 and 1988, the Federal Communications Commission (FCC) opened two major dockets to solicit and consider alternatives to rate of return regulation.[15] In late 1986, the National Telecommunications and Information Administration (NTIA), which is an Executive branch agency charged with developing national telecommunications policy, solicited comments about the efficacy of rate of return regulation as applied to the telecommunications industry,[16] and subsequently released a report that recommended substantial changes in how the industry is regulated.[17]

Between 1984 and 1987, at least 19 states enacted telecommunications "deregulation" legislation that generally directed and empowered state regulators to seek alternatives to traditional regulation for the telecommunications industry.[18] In a survey of state telecommunications legislation, the Michigan Public Service Commission reported that between 1987 and mid-1988, 34 regulatory reform or "deregulation" bills were introduced in state legislatures and 17 were enacted.[19] Another recent survey reported that 12 state telecommunications laws were enacted in the first half of 1988 that substantially modified the application of traditional rate of return regulation.[20] Scores of state public utility commissions have opened dockets to consider alternatives to rate of return regulation for telecommunications regulation.

---

[14] See, e.g., Sedgley & Cody, The Effect of Deregulation on AT&T's Pricing, 120 PUB. UTIL. FORT. 34 (Nov. 26, 1987); Mathios & Rogers, The Impact of Alternative Forms of State Regulation of AT&T on Direct Dialing Long Distance Telephone Rates, FTC Working Paper series No. 159 (Dec. 1987); Frentrup, The Effect of Competition and Regulation on AT&T's Intrastate Toll Prices, and of Competition on Bell Operating Company IntraLATA Toll Prices, FCC Common Carrier Bureau working paper (June 22, 1988); Perl, Welfare Consequences of Competition in Telecommunications, paper presented at 7th International Conference of the International Telecommunications Society (June 29 - July 1, 1988); and, B. Kahn, Price Caps vs. Cost of Service Regulation: Evidence on Price Performance from the Intrastate Toll Market, paper presented at Rutgers University Advanced Workshop in Public Utility Economics and Regulation, Western Conference (July 6-8, 1988).

[15] Decreased Regulation of Certain Basic Telecommunications Services: Notice of Proposed Rulemaking, CC Docket 86-421, 2 FCC Rcd 645 (Jan. 9, 1987); and, Policy and Rules Concerning Rates for Dominant Carriers: Notice of Proposed Rule Making, CC Docket 87-313, 2 FCC Rcd. 5208 (1987), ______: Further Notice of Proposed Rulemaking, ___ FCC Rcd. ____ (May 23, 1988).

[16] Comprehensive Review of Rate of Return Regulation of the Telecommunications Industry: Notice and Request for Comments, NTIA docket No. 61091-6191, 51 Fed. Reg. 36,837 (Oct. 16, 1987).

[17] NTIA, U.S. DEPT. OF COMMERCE, NTIA REGULATORY ALTERNATIVES REPORT (July 1987) (hereafter cited as NTIA REGULATORY ALTERNATIVES REPORT).

[18] See 1985 Ariz. Sess. Laws H.B. 2149; 1987 Colo. Sess. Laws H.B. 1336, 1985 Conn. Acts 85-187 & 1987 Conn. Acts 87-415, 1985 Ill. Laws H.B. 1814; 1985 Ind. Acts S. 530; 1983 Iowa Acts H.F. 312; 1986 Mich. Pub. Acts H.B. 5158; 1987 Minn. Laws S.F. 677; 1987 Mo. Laws H.B. 360; 1985 Mont. Laws H.B. 577; 1986 Neb. Laws L.B. 835; 1985 N.M. Laws H.B. 271 & 559; 1985 N.D. Sess. Laws S.B. 2201; 1985 Or. Laws H.B. 2200; 1987 Tex. Gen. Laws S.B. 444 & 229; 1984 Utah Laws S.B. 102; 1987 Vt. Acts S. 114; 1985 Wash. Laws S.B. 3305; and 1986 Wis. Laws S.B. 318.

[19] COMM. DAILY 6 (Aug. 2, 1988). Most of the deregulation bills concerned social contract alternatives to rate of return regulation, deregulation of competitive telecommunications services, deregulation of small telephone companies, and deregulation of cellular services.

[20] Deregulation, Administrative Changes Lead Topics of 1988 State Phone Laws, 6 STATE TELEPHONE REGULATION REPORT 1, 4 (July 28, 1988).

There is a trend toward regulatory change in other countries as well. The Japanese have recently enacted major revisions to their telecommunications regulation that have reduced the degree of government oversight.[21] European countries are also moving away from pervasive regulation and towards the provision of telecommunications services in a private, competitive marketplace.[22]

This paper contributes to this investigation by surveying and analyzing the major alternatives to rate of return regulation.[23] It consists of three major sections:

I. *Evaluation Criteria* -- describes the criteria that are relevant to evaluating alternatives to rate of return regulation for the telecommunications industry;

II. *Minor Regulatory Changes* -- describes alternatives that retain traditional regulation's basic revenue requirements structure, but involve minor modifications to the calculation and use of revenue requirements; and,

III. *Major Regulatory Changes* -- describes alternatives that abandon or radically alter rate of return regulation's basic revenue requirements structure.

The concepts presented in this paper are summarized in Table 1, *Evaluation Criteria*, and Table 2, *Alternatives to Rate of Return Regulation in the Telecommunications Industry.*

I. EVALUATION CRITERIA

Eight broad regulatory goals are relevant to an evaluation of alternatives to rate of return regulation in the telecommunications industry:[1]

---

[21] For a survey of Japanese telecommunications regulation, see Harris, Telecommunications Policy in Japan: Lessons for the U.S., paper presented at Rutgers University Advanced Workshop in Public Utility Economics and Regulation, Western Conference (July 6-8, 1988).

[22] For a comparative survey of Western telecommunications regulatory practices, see Mansell, Telecommunications Network-Based Services: A Comparative Analysis of Regulatory Frameworks, paper presented at 7th International Conference of the International Telecommunications Society, Cambridge, MA (June 28 - July 1, 1988).

[23] Other surveys of alternatives include, BONBRIGHT, 547-646; Brown, Alternatives to Rate of Return Regulation, presented to Communications Subcommittee, National Association of Regulatory Utility Commissioners conference, San Francisco, CA (July 20, 1987); CREW & KLEINDORFER, 146-168; and, Eberle & Williamson, Deregulation of Telecommunications at the State Level: Managing a Transition, 122 PUB. UTIL. FORT. 20 (Sept. 1, 1988). An interesting analogy between rate of return alternatives and the alternatives employed to regulate weapons systems contracts is in Kovacic, The Expanding Use of Competition Strategies in the Acquisition of Major Weapon Systems, paper presented at Rutgers University Advanced Workshop in Public Utility Economics and Regulation, Western Conference (July 6-8, 1988).

[1] This list roughly corresponds with the seven characteristics of effective regulation recently listed by the NTIA: Reasonable rates, cost minimization, innovation, price flexibility, administrative efficiency, adequate returns, and fairness. NTIA REGULATORY ALTERNATIVES REPORT, 11-12.

1. *Protect Customers.* An alternative should leave customers at least as well off as they are under existing institutions, and possibly better off. Protection of universal service and promotion of service to selected customer classes (e.g. low income, elderly and disabled subscribers) are included under this criteria.
2. *Protect Investors.* An alternative should leave investors at least as well off as they are under existing institutions, and possibly better off.
3. *Promote Fairness.* An alternative should be fair to customers, investors, competitors and other affected parties.
4. *Promote Competition.* An alternative should promote competition and reduce incentives for dominant firms to engage in anticompetitive acts, such as monopoly pricing (excessive rates or returns) or predatory pricing (rates designed to injure competitors).
5. *Enhance Economic Efficiency.* An alternative should enhance economic efficiency by driving service prices closer to the economic costs of providing those services.
6. *Enhance Administrative Efficiency.* An alternative should enhance the administrative efficiency of existing institutions.
7. *Promote Innovation/Maintain Service Quality.* An alternative should increase incentives for firms to adopt innovative means of providing service, while maintaining existing high levels of service quality.
8. *Implementation.* An alternative should be politically and legally feasible.

Protecting telecommunications consumers is probably the most important telecommunications regulatory goal. Protection of consumer interests is often cast as having three elements -- protection of existing high levels of telephone subscribership;[2] provision of service at the lowest possible rates; and, promoting service to disadvantaged groups (e.g. low income, elderly, and disabled subscribers).[3] Obviously, many of these consumer interest goals are also related to basic concepts of fairness.

Protection of consumer interests is often perceived of as balanced against investor interests. Constitutional constraints on regulation require that investors receive *just* compensation for property devoted to public service and that Constitutional *right* often is balanced against consumers' interests in reasonable rates.[4]

---

[2] Protection of universal service is specified by statute in some states. For example, a New York statute requires, as part of a major rate change, the expected impact that the change will have on universal service. N.Y. Public Service Law § 92(4) (Supp. McKinney 1987).

[3] See generally, Annot., Validity of Preferential Rates for Elderly or Low Income Persons, 29 ALR 4th 606. A number of states have created "universal service funds" or lifeline programs provide subsidies for low income subscribers. The FCC has implemented a low income lifeline program to provide low income subscribers with assistance in paying federal access charges and a "Link-up America" program to provide funds for low income connection charges. See generally, MTS & WATS Market Structure: Recommended Decision and Order, CC Dockets 78-72 & 80-286, 2 FCC Rcd 2324 at par. 71 (Mar. 31, 1987), adopted by the Commission, 2 FCC Rcd 2953 (May 19, 1987). Also, federal statutes and some states statutes have provisions which promote telephone service to disabled persons. See, e.g., 47 U.S.C. § 610(a).

Promotion of fairness or equity is a basic goal of utility regulation. Institutional evidence of the pervasive nature of fairness as a telecommunications regulatory objective is found in statutory standards such as *just and reasonable* rates, the proscription against offerings that grant an *undue preference or advantage*, general proscriptions against discrimination, the concern for disadvantaged groups who might be adversely affected by regulatory changes, and the adherence to administrative processes to guarantee that affected parties have a right to be heard.[5] Also, the perception that telecommunications services are a "right" that should be available to all is rooted in equity considerations.[6]

Promotion of competition is an increasingly important telecommunications regulatory goal, as well as a means used to achieve regulatory ends (promotion of economic efficiency, enhance innovation incentives, etc.). Some state statutes explicitly recognize "promotion of competition" as a telecommunications regulatory goal.[7] The FCC has often stated that promotion of competition is one of its basic telecommunications regulatory objectives. Also, the pervasive coverage of antitrust laws, extending even to regulated industries with limited exceptions, evidences a national pro-competitive policy.[8]

Economic efficiency is derived from the traditional regulatory interest in "cost-based" rates. Indeed, one of the major rationales for traditional rate of return regulation is that regulated firms should recover no more than the total costs (revenue requirements) associated with providing service. Similarly, rate design issues often turn on economic efficiency issues.[9] While economic efficiency and promotion of competition are theoretically linked, it should be noted that they are distinct -- an alternative that promotes economic efficiency (competition) does not necessarily promote competition (economic efficiency).

---

[4] For a more detailed analysis or these constraints as applied to the telecommunications industry, see Levitz, Loosening the Ties that Bind: Regulating the Interexchange Services Market for the 1990's, 2 FCC Rcd 1495 (Mar. 12, 1987).

[5] Promotion of fairness has a huge, and growing literature. For surveys of equity concepts in public utility economics, *see* E.E. ZAJAC, FAIRNESS OR EFFICIENCY: AN INTRODUCTION TO PUBLIC UTILITY ECONOMICS; Zajac, *Perceived Economic Justice: The Example of Public Utility Regulation*, in COST ALLOCATION: METHODS, PRINCIPLES, APPLICATIONS, 119 (Young, ed. 1985); and, Sievers, But is it Fair? Economic Fairness Standards Applied to Federal Telecommunications Access Charges, presented at Rutgers University Advanced Workshop in Regulation and Public Utility Management (April 8, 1988).

[6] *See, e.g.*, Zajac, *Is Telephone Service an Economic Right?* in ENERGY AND COMMUNICATIONS IN TRANSITION 94 (Proceedings of the Institute of Public Utilities, 1981).

[7] For example, Mo. Rev. Stat. § 392.530(6) (1987) directs the Missouri Public Service Commission to "[a]llow full and fair competition to function as a substitute for regulation when consistent with the protection of ratepayers and otherwise consistent with the public interest."

[8] *See generally*, Areeda, *Antitrust Laws and Public Utility Laws*, 3 BELL J. ECON. & MGT. 42 (1972); SCHWARTZ *et al.*, 830-959; and, Sievers, The Transition to Antitrust Law Regulation in the Telecommunication Industry: A Comparison of Public Utility Law Regulation and Antitrust Law Regulation, presented to Communications Subcommittee, National Association of Public Utility Commissioners, San Francisco conference July 19-20, 1987.

[9] For surveys of rate design methodologies, *see*, SCHWARTZ, *et al.*, 659-829; PHILLIPS, 379-462; and BONBRIGHT, 377-546.

Administrative efficiency can be traced to the sections of regulatory statutes devoted to administrative and organizational matters. At the federal level, the FCC is directed to "conduct its proceedings in such a manner as will best conduce to the proper dispatch of business and to the ends of justice."[10] Similarly, most state utility regulation statutes devote considerable language to administrative matters. Jurisdictional issues are also related to administrative efficiency -- what mix of federal and state regulation is appropriate for services and facilities that have both inter and intrastate characteristics?

Promotion of innovation and maintenance of service quality is derived from technological innovation objectives in telecommunications. For example, federal telecommunications regulatory statutes[11] and some state statutes[12] include promotion of technological innovation as an explicit regulatory objective. Protection of service quality is an important regulatory objective that is bundled together with innovation. Obviously, an alternative that creates incentives to adopt a cost cutting innovation by reducing service quality would not be acceptable.

An alternative should also be legally and technically feasible to be seriously considered. It does little good to adopt an alternative only to have it struck down by reviewing Courts as unconstitutional or inconsistent with statutory obligations.[13]

While it is not possible to distinctly apply each of these criteria to all of the alternatives that follow, an attempt was made to use these criteria to make a broad evaluation of the each of the alternatives.

II. MINOR REGULATORY CHANGES

The sections that follow discuss alternatives that retain a basic revenue requirement structure, but incorporate relatively minor modifications to that structure. The alternatives include:

A. *Negotiated/Abbreviated Rate Proceedings.* Elements of revenue requirements or rates are determined outside of a formal regulatory environment, and regulators are asked for approval in an abbreviated proceeding. Two types of outside proceedings are considered: (1) social contract proceedings; and, (2) "fast track" proceedings which include streamlined regulation for competitive services and automatic adjustment mechanisms similar to the electric utility industry's fuel adjustment clauses.

---

[10] 47 U.S.C. § 154(j).

[11] 47 U.S.C. § 157(a). "It shall be the policy of the United States to encourage the provision of new technologies and service to the public. Any person or party (other than the Commission) who opposes a new technology or service proposed to be permitted under this chapter shall have the burden to demonstrate that such proposal is inconsistent with the public interest." Also, 47 U.S.C. § 218 directs the FCC to "keep itself informed ... as to technical developments and improvements in wire and radio communication and radio transmission of energy to the end that the benefits of new inventions and developments may be made available to the people of the United States."

[12] For example, Texas' utility laws direct the Texas Public Utilities Commission to "balance the public interest in a technologically advanced telecommunications system providing a wide range of new and innovative services with traditional regulatory concerns." Tex. Rev. Civ. Stat. Ann. art. 1446c, § 18(g) (Supp. Vernon 1987).

[13] See, e.g., Farmers Union Central Exchange, Inc. et al. v. FERC, 734 F.2d 1486 (D.C. Cir. 1984), cert. den. Williams Pipe Line Co. v. Farmers Union Central Exchange, Inc., 469 U.S. 1034 (1984) where the court struck down a FERC rulemaking decision that radically altered rate regulation for natural gas pipelines because the FERC failed to relate its regulatory changes to its statutory obligations.

B. *Rate of Return Bands/Incentive Regulation.* A rate of return band is set such that the lower limit is the allowed rate of return used in revenue requirement calculations. Earnings between the lower and upper limits are presumed due to efficiency enhancements and either kept by the firm or shared with customers in the form of temporary rate reductions. Earnings above the upper limit are refunded to customers. A variant of this is no upper limit on earnings, but a graduated sharing where the consumers' share increases as earnings increase.

C. *Price Bands.* Regulators set a price band such that regulated firms are allowed to alter their prices within the band with limited, or no regulatory oversight. A variant of this is to set only an upper limit on price changes, but no lower limit.

D. *Monopoly Services Regulation.* Rate of return regulation is selectively applied to only those services (or firms) that are variously classified as natural monopoly services, "core" services, or common carrier services. A variant of this is prospective *deregulation* of new services, since new services are arguably non-basic, non-core, and non-monopoly services.

A. Abbreviated Rate Proceedings

Two types of abbreviated rate proceedings are popular among utility regulators: social contract proceedings, and "fast track" proceedings that include streamlined regulation of competitive services and mechanisms similar to fuel adjustment clauses in the electric utility industry. Under both alternatives, the elements of revenue requirements and/or individual rates are calculated outside of general rate cases, and the Commission's formal role is reduced to one of evaluating the proposal in an abbreviated proceeding.

Ratemaking outside of formal regulated proceedings, such as negotiated rate settlements,[1] have long been a part of utility regulation. For example, in the trucking industry, rates are set through collective ratemaking where industry "rate bureaus" determine trucking rates and submit the rate to regulators for approval.[2] Prior to the late 1960's, AT&T and the FCC often engaged in informal negotiations to set rates.[3] State Commission staffs often negotiate with utilities about regulatory matters and, when agreement is possible, present a joint proposal to the Commission.

---

[1] See, e.g., O'Leary, Negotiated Settlements in Utility Regulation, 118 PUB. UTIL. FORT. 11 (Aug. 11, 1986); and, Susskind & McMahon, The Theory and Practice of Negotiated Rulemaking, 3 YALE J. on REGULATION 133 (1985).

[2] See generally, MOTOR CARRIER RATEMAKING STUDY COMMISSION, COLLECTIVE RATEMAKING IN THE TRUCKING INDUSTRY (June 1983); and the Court's description in, Southern Motor Carriers Rate Conference v. United States, 471 U.S. 48 (1985) [holding that ratemaking by rate bureaus is exempt from the application of the antitrust laws when supervised by state regulators and part of a "clearly articulated" state policy to displace competition].

[3] See AT&T, Docket 16258, 2 FCC 2d 173, 177 (1965) as cited in Policy and Rules Concerning Rates for Dominant Carriers: Notice of Proposed Rule Making, CC Docket 87-313, 2 FCC Rcd 5208 (1987). The FCC referred to its informal rate negotiations with AT&T as "continuing surveillance." Also see, NTIA REGULATORY ALTERNATIVES REPORT at 7-8.

1. *Social Contract*

Social contract proceedings do not necessarily modify the basic structure of rate proceedings, but allow regulators and utilities to negotiate and arrive at mutually agreeable rates and revenues outside of formal rate cases. In other words, regulators and utilities are freed from formal rate proceedings and allowed to enter into a "social contract" that, when accepted by the Commission, has the same force of law as would rates derived from a traditional rate case.[4] Some authors describe a social contract alternative as negotiations between a telephone company and the regulator where the telephone company agrees to limit rate increases for some services, usually basic local service, in exchange for more pricing freedom on other services or relaxation of the allowed rate of return.[5] The unifying element of social contract proceedings, however, is the negotiating and contracting *process that takes place outside of formal proceedings*, not the terms ultimately adopted.

The phrase "social contract" was popularized by Vermont Commissioner V. Louise McCarren.[6] The idea behind the phrase was that traditional regulation is an expression of a "social contract" between telecommunications utilities and society (ratepayers), where society agreed to limit market entry in exchange for the utility's promise to limit its prices and profits and to engage in a socially beneficial cross-subsidization of local service rates. The introduction of competition, however, has made maintenance of this local service subsidy more difficult, and has necessitated the creation of a new "social contract."

The recent order of the Vermont Public Service Board is an example of "social contract" regulation.[7] New England Telephone and the Vermont Department of Public Service drafted a "Vermont Telecommunications Agreement" (VTA), a "contract" that covered a wide range of regulatory issues, including local service rates, revenue regulation, and access charge regulation. Interested parties were then allowed to comment on the VTA. The Vermont Public Service Board took those comments under consideration and issued an order generally approving the agreement with certain modifications.

A number of other states have also engaged in various forms of

---

[4] See the description of social contract alternatives in Comprehensive Review of Rate of Return Regulation of the Telecommunications Industry: Notice and Request for Comments, National Telecommunications and Information Administration docket no. 61091-6191, 51 Fed. Reg. 36,837 (Oct. 16, 1986).

[5] See, e.g., Landau, Social Contract Regulation Is a Bad Bargain for Ratepayers, 120 PUB. UTIL. FORT. 21 (July 9, 1987); and, "Social Contract" Plans to Control Local Telephone Rate Increases Ignore New Deflationary Trends in Telco Revenue Requirements, 11 TRENDS IN COMMUNICATION REGULATION 3 (Nov. 1986). Also see Lay, A Pragmatic Look at Social Contracts and Price Caps, 4 TELEMATICS 1 (Oct. 1987) who describes social contract regulation as a form of price cap.

[6] See, e.g., McCarren & Dworkin, State Regulators, State Competitors, in THE NEW TELECOMMUNICATIONS ERA AFTER THE AT&T DIVESTITURE: THE TRANSITION TO FULL COMPETITION 147 (H. E. Forrest & R. E. Wiley, eds. 1985).

[7] Investigation of proposed "Vermont Telecommunications Agreement": Order, Docket No. 5253 (July 12, 1988).

"social contract" regulation.[8] For example, in February 1987, the New York Public Service Commission approved an agreement between New York Telephone, New York State, and the New York State Consumer Protection Board that reduced rates by $100 million in 1987, provided for further reductions of $300 to $400 million by 1990, and extended an existing moratorium on general rate increases through 1991.[9]

2. *Fast Track Proceedings*

Fast track proceedings are another form of abbreviated rate proceeding. Two types of *Fast Track* mechanisms are employed in utility regulation: (1) streamlined tariff approval for competitive services; and, (2) mechanisms similar to fuel adjustment clauses in the electric utility industry.

Streamlined tariff review is offered by both federal and state regulators. For example, in early 1987, the FCC proposed to allow packet services and contract services offered by otherwise dominant carriers to be subject to streamlined tariff review.[10] Similarly, some states' regulators apply streamlined tariff review to competitive services, which generally allow a tariff to go into effect a short time after it is filed without extensive regulatory analysis. For example, Ohio has applied streamlined review of competitive services since 1985.[11] Streamlined review of competitive services is specified by statute in a number of states.[12]

In the electric utility industry, fuel adjustment clauses are used to deal with rapidly changing fuel prices.[13] A typical fuel adjustment clause requires the utility to set its rates based on its forecast of fuel prices. To the extent that the utility's fuel forecast is wrong, the shortfall (overage) of actual revenues is added to (subtracted from) a non-earning accumulating account. At any time, the utility may petition the Commission for a temporary increase (decrease) in rates to recover (refund) the shortfall (overage) in the account. Generally, approval of the temporary rate change is automatic and immediately implemented, unless someone protests the increase (decrease) and proves that it is unjustified. This is in contrast to the burden of proof in a typical rate proceedings where the utility must prove that its rate change is justified.

---

[8] For the trade press description of various social contract plans, see, e.g., Michigan Bell Seeks Social Contract; Similar Plans See Action in Maryland, Vermont, and New Mexico, 6 STATE TELEPHONE REGULATION REPORT 1 (Jan. 14, 1988); Latest "Social Contract" Proposal Offer by BOCs in New Mexico, Maryland, 5 STATE TELEPHONE REGULATION REPORT 10 (Oct. 8, 1987); and, Consumer Advocates Oppose Social Contract Regulation of Telco Rates, 5 STATE TELEPHONE REGULATION REPORT 1 (Nov. 19, 1987).

[9] AMERICAN BAR ASSOCIATION, 1987 ANNUAL REPORT, SECTION OF PUBLIC UTILITY LAW, 114 (1987).

[10] Decreased Regulation of Certain Basic Telecommunications Services: Notice of Proposed Rulemaking, CC Docket 86-421, 2 FCC Rcd. 645 (Jan. 9, 1987).

[11] See, e.g., In the Matter of the Commission's Investigation into the Regulatory Framework for Competitive Telecommunication Services in Ohio, Case No. 84-944-TP-COI (April 9, 1985) and Case No. 86-1144-TP-COI (Aug. 2, 1988).

[12] See, e.g., Colo. Rev. Stat. § 40-15-301 et seq. (Supp. 1988); Fla. Stat. Ann. § 364.337 (Supp. 1988); 1988 Ga. Laws HB-1234; Ill. Ann. Stat. ch. 111 2/3 §§ 13-502 et seq. (Supp. 1988); Ind. Code Ann. 8-1-2.6-1 et seq. (Supp. 1988); Iowa Code Ann. § 476.1 (Supp. 1986); Minn. Stat. Ann. sec 237.60 (Supp. 1988); Mo. Rev. Stat. §§ 392.010 et seq. (Supp. 1988); Mont. Code Ann. 69-3-807 (1985); N. M. Stat. Ann. §§ 63-9A-1 et seq. (Supp. 1985); N. D. Cent. Code § 49-21-02.1 (Repl. 9B, Supp. 1987); Tex. Rev. Civ. Stat., art. 1446c, § 18(a) et seq. (Supp. Vernon 1987); Utah Code Ann. § 54-8b-1 et seq. (Supp. 1985); Va. Code § 56-481.1 (Repl. 1985); Wis. Stat. § 196.195 (Supp. 1987); and, Wash. Rev. Code Ann. §§ 80.36.300-340 (Supp. 1988).

[13] For a description of fuel adjustment plans and cases, see Annot., Validity of "Fuel Adjustment" or Similar Clauses Authorizing Electric Utility to Pass on Increased Cost of Fuel to Its Customers, 83 A.L.R. 3d 933.

A similar institution can be carried into the telecommunications industry. Regulators could set a target return according to its existing rules and allow firms to accumulate earnings errors between the actual and target return. If the return was above (below) the target return, then customers would receive refunds (pay a surcharge). In order to avoid "cost plus" incentives, the accumulating account could be structured to accumulate only earnings errors due to unanticipated market risks.[14]

The California Public Utilities Commission's attrition mechanism is an example of this alternative. Generally, telecommunications utilities are subject to general rate cases every three years and an attrition adjustment is made in the off year. The attrition proceeding focuses on just total earnings and not the rate design issues that would be a part of a general rate case. The use of the attrition mechanism is not limited to utilities to recover unanticipated costs. For example, the Division of Ratepayer Advocates recently petitioned the California Commission for an attrition adjustment for AT&T based on revenue changes attributable to exogenous access charge reductions, increased toll traffic due to reduced rates, tax law changes, and increased productivity.[15]

3. *Analysis*

Both of the *Abbreviated Rate Proceeding* alternatives are administratively more efficient than traditional rate of return regulation. Both decrease the regulatory burden associated with rate regulation in traditional, formal administrative proceedings. Since both alternatives rely on individual firm data and negotiations, both arguably enhance economic efficiency by moving regulated prices closer to individual firms' economic costs. Also, since the *Fast Track* alternative reduces the firm's exposure to unanticipated risks that it would otherwise not be compensated for, it arguably enhances the fairness of traditional regulation. Also, the *Fast Track* alternative allows customers to realize cost reductions before they reach a level that justifies a general rate case.

On the negative side, when the *Fast Track* alternative is used to reduce revenues between rate cases, such as the above described use of the California Commission's attrition mechanism, innovation and cost reduction incentives are reduced. A firm will be less likely to engage in innovative or cost reducing research and development if it knows that its cost savings will be immediately flowed through its rates.[16] Thus, it is questionable whether the administrative and economic efficiency savings outweigh the technical efficiency costs enough to warrant adopting the *Fast Track* alternative.

---

[14] Market risks are the non-firm specific risks associated with a firm's required rate of return. Generally, market risks are composed of expected inflation and market interest rate volatility. For example, the accumulating account could contain errors due to incorrect inflation and market interest rate forecasts. The rationale for this alternative is that a firm should not be penalized for changes in costs and risks it cannot control. Regulation imposes constraints on the firm's ability to maximize profits, so it is unfair to make the firm bear the burden of losses due to unanticipated and uncontrollable costs and risks.

[15] In the Matter of the Application of AT&T Communications of California, Inc., a Corporation, for Authority to Increase Certain Intrastate Rates and Charges Applicable to Telecommunication Service Furnished Within the State of California (U 5002 C): Motion to Reopen Proceedings to Provide for an Attrition Review, Application No. 85-11-029 (Aug. 4, 1988).

The often stated purpose of allowing a service to be subject to streamlined regulation is to allow a regulated firm the flexibility needed to respond to competition. However, the *Abbreviated Rate Proceeding* alternatives can be used to achieve anticompetitive objectives. Since competitors' interests are not typically represented either in negotiations between a utility commission staff or when a utility makes an attrition adjustment, both alternatives are subject to abuse by a utility that seeks to adjust its rates to achieve anticompetitive results. For example, in California AT&T recently requested a six month temporary rate reduction to refund excess earnings due to one-time revenue requirement adjustments rather than refund the excesses in a lump sum payment to ratepayers.[17] US Sprint argued that such temporary rate reductions have an anticompetitive impact because they create the illusion that AT&T's rates are lower than its competitors' rates. The final resolution of the matter is still pending before the California Commission.

The *Social Contract* alternative also presents the danger that regulated firms will engage in anticompetitive acts in the service markets where they gain pricing freedom. Since the *Social Contract* alternative often allows price changes in non-basic services with little or no supporting cost information, it is difficult, if not impossible, to tell whether the price changes are the result of legitimate responses to competition or anticompetitive changes. The problem is exacerbated if a regulated firm is granted pricing freedom in non-competitive markets, or where competitors must rely on services provided by the regulated firm in order to compete with it (i.e. IXCs must obtain access services offered through LEC local facilities and must compete with LECs in intrastate, intraLATA toll markets).

The *Social Contract* alternative usually eliminates the cost studies, price-cost relationships, and cost allocation processes of traditional rate of return regulation that are usually thought to be necessary to detect and prevent excessive and/or anticompetitive pricing.[18] Thus, the *Social*

---

[16] The impact of regulation on innovation is not theoretically well developed. Broadly speaking, there are two schools of economic thought: (1) the Schumpeter school holds that innovation can be promoted only when the firm is allowed to realize some monopoly profits on its innovation; and, (2) the Arrow school holds that innovation is greater in competitive than non-competitive market structures. A good survey of the literature and a synthesis of these two schools is in Arthur D. Little, Inc., The Relationship Between Market Structure and the Innovative Process (Jan. 1976), reprinted in Domestic Telecommunications Common Carrier Policies: Hearings Before the Subcommittee on Communications of the Senate, 95th Cong., 1st Sess. 309 at 319 (1977). Surveys of the regulation-innovation literature are in Khanna, Economic Regulation and Technological Change: A Review of the Literature, 109 PUB. UTIL. FORT. 35 (Jan. 21, 1982); and, F.W. SCHERER, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE, 407-438 (2d 1980). Some empirical tests of the impact of regulation on innovation are in Smith, The Implications of Regulation for Induced Technical Change, 5 BELL J. ECON. & MGT. (1974); Okuguchi, The Implications of Regulation for Induced Technical Change: Comment, 6 BELL J. ECON. & MGT. 703 (1975); Denny, Fuss, Everson & Waverman, Estimating the Effects of Diffusion of Technological Innovations in Telecommunications: The Production Structure of Bell Canada, 14 CAN. J. ECON. 24 (1981); and, Nelson, Regulation, Capital Vintage, and Technical Change in the Electric Utility Industry, 66 REV. ECON. & STAT. 59 (1984).

[17] AT&T Advice Letter No. 97 (May 20, 1988).

[18] The cost allocation and pricing methods of traditional regulation are extensively criticized. See, e.g., Baumol & Walton, Full Costing, Competition and Regulatory Practice, 82 YALE L. J. 639 (1973); Braeutigam, An Analysis of Fully Distributed Cost Pricing in Regulated Industries, 11 BELL J. ECON. & MGT. 182 (1980); Baumol, Koehn &

Contract alternative represents a major change in perceptions about the efficacy of these traditional rate of return institutions -- under rate of return regulation, cost justification is presumed to be necessary to prevent abuse of monopoly power, whereas under the *Social Contract* alternative, these institutions are presumed to be unnecessary, or at least replaceable by a negotiation process.

A danger of the *Social Contract* alternative is that regulators and regulated firms enter into negotiations with very different levels of information about the regulated firm. Without data, regulators do not, and cannot, know a regulated firm's earnings or what its business plans are. Regulators can protect consumers by forcing the firm to agree to "cap" rates at existing levels, but they cannot know whether lower rates would have been realized under rate of return regulation. For example, if a local exchange carrier is currently earning in excess of its allowed rate of return, traditional regulation might force the firm to disgorge these excess profits in the form of rate decreases. However, if the firm enters into a "social contract" with its regulator before the regulator discovers the excess earnings and conceals its earnings in the negotiation process, ratepayers are made worse off by *Social Contract* regulation.

The *Social Contract* alternative may or may not enhance fairness. To the extent that the negotiations are throught of as voluntary and the interests of all parties are fully represented, then the *Social Contract* alternative is arguably fair. However, if the negotiations exclude the interests of affected parties or if there are significant asymmetries of information in the negotiation process, then the result may not be fair.

The *Social Contract* alternative is probably not feasible in its pure form for the FCC, but is feasible for state telecommunications regulators. The FCC's implementation of a *Social Contract* alternative is limited by the number of telecommunications carriers -- individualized negotiations are certainly not possible with each of the 1400 exchange carriers that the FCC oversees. Given the FCC's resources, such negotiations may not be possible even if the FCC limits its negotiations to the Bell Operating Companies (BOCs) and AT&T.

B. Rate of Return Bands/Incentive Regulation

Under this alternative, a regulator would not set a single allowed rate of return, but would set a rate of return band. For example, a regulator might set the lower limit of the band at 12% and the upper limit at 14%. The lower limit could be used in setting revenue requirements and in jurisdictional separations processes, so it would not change the way that revenue requirements are calculated. Firms would be free to earn below the lower limit and would refund earnings in excess of the upper limit. Earnings between the upper and lower limits of the band are presumed due to increases in the firm's efficiency and would be shared between the firm and its customers so as to provide the firm with cost-cutting efficiency

---

Willig, How Arbitrary is Arbitrary? -- Or, Toward the Deserved Demise of Full Cost Allocation, 120 PUB. UTIL. FORT. 16 (Sept. 3, 1987); Temin & Peters, Cross-Subsidization in the Telephone Network, 21 WILLAMETTE L. REV. 199 (1985); ______, Is History Stranger Than Theory? The Origin of Telephone Separations, 75 AM. ECON. REV. 324 (1985); HUBER, THE GEODESIC NETWORK: 1987 REPORT ON COMPETITION IN THE TELEPHONE INDUSTRY, 3.54-3.55, U.S. Department of Justice report (Jan. 1987); MCI Telecommunications Inc. v. AT&T, 708 F. 2d 1081, 1116-1117 (7th Cir. 1983), cert. den., 464 U.S. 891 (1983), appeal after remand, 748 F.2d 799 (7th Cir. 1984);1 KAHN, 150-158; BONBRIGHT, 408-513; and, Larson, Costs, Allocations, and Regulatory Issues, 5 TELEMATICS 11 (April 1988).

incentives. As an alternative, no upper limit might be specified, but the consumers' share in earnings could increase with earnings.

In lieu of *ad hoc* changes in consumer and regulated firm shares in excess earnings, measures of managerial efficiency could be used to determine the firm's share in earnings. For example, an adjustment could be made to the proportion of excess earnings retained by the firm according to objective measures of managerial efficiency.[19] Since some states currently tie portions of utility returns to managerial efficiency,[20] such an efficiency based sharing arrangement is arguably consistent with current practice.

The rationale for this alternative is that it creates economic incentives for efficiency enhancements. It also recognizes that calculation of rate of return is largely a subjective process; it is not realistic to believe that a single rate of return for all exchange carriers can be determined with eight digits of precision. Rather, the rate of return is somewhere within a "zone of reasonableness."[21]

There are many examples of rate of return banding. In 1986, the Alabama Public Service Commission approved a Rate Stabilization and Equalization plan for South Central Bell which specified a rate of return band between 11.65% and 12.3%.[22] New Jersey specifies a 3% band around AT&T's target rate of return. In February 1987, as part of its "social contract", New York Telephone agreed to share half of its earnings above its current 14% return on equity with its ratepayers in the form of rate reductions. In June 1987, the Connecticut Department of Public Utility Control approved a negotiated agreement between the Connecticut Consumer Counsel and Southern New England Telephone (SNET) where SNET agreed to freeze basic service rates until 1989 and to share half of its earnings between 13.5% and 14.3% with its customers. In July 1987, Wisconsin Bell was ordered to share earnings between 14.0% and 15.5%, and agreed not to seek a rate increase unless its earnings fell below 12.5%.[23]

1. *Analysis*

Rate of Return Bands/Incentive Regulation reduces the regulatory burden of traditional regulation since firms and regulators do not have to worry about earnings in excess of the allowed return. It is simpler than existing regulation since regulators do not have to engage in artificial, highly abstract ruminations about *the* allowed rate of return, but merely have to specify the range within which it falls. The alternative also enhances innovation since it increases incentives for firms to adopt cost-cutting measures and keep a portion of the profits generated thereby.

---

[19] For surveys of managerial standards in regulated industries, see Burkhart, Objective Standards of Management Efficiency Have Arrived for Regulated Utilities, 118 PUB. UTIL. FORT. 48 (Nov. 27, 1986); Joskow & Schmalensee, Incentive Regulation for Electric Utilities, 4 YALE J. on REGULATION 1 (1986); Sloane, The Rate of Return on Common as the Overall Measure of Utility Management Performance, 115 PUB. UTIL. FORT. 37 (Feb. 7, 1985); and, Hall, Performance Measurement Techniques: A Review, 116 PUB. UTIL. FORT. 25 (Aug. 22, 1985).

[20] See, e.g., Kan. Stat. Ann. §§ 66-128a-128g (1985).

[21] This statement reflects the FCC's practice of setting a single allowed rate of return for all local exchange carriers. It justifies a single allowed rate of return by noting that the capital structures and risks of local exchange carriers are not materially different between firms and by invoking the doctrine of "administrative inconvenience." For an analysis of this issue, see Lavey, Doctrine of Administrative Inconvenience at the Federal Communications Commission, 17 LOYOLA U. L. J. 617, 624-631 (1986).

[22] Re South Central Bell Telephone Company, Docket 19983, 78 PUR 4th 299 (Ala. PSC Nov. 3, 1986).

[23] AMERICAN BAR ASSOCIATION, 1987 ANNUAL REPORT, SECTION OF PUBLIC UTILITY LAW, 114 (1987).

The incentives to adopt cost-cutting, profit increasing innovations and the decrease in administrative costs associated with the *Rate of Return Bands/Incentive Regulation* alternative could result in lower rates for customers simply because of reduced costs. Similarly, improved profits translate into higher returns for investors. Investor uncertainty will also be reduced as there will be less regulatory risk associated with the determination of the allowed rate of return. Also, from an investor's viewpoint, the allowed rate of return will be less likely to constrain the earnings of the regulated firm.

Rate of Return Bands/Incentive Regulation is legally feasible. Since courts and regulators often refer to the rate of return as falling within a "zone of reasonableness",[24] a rate of return band naturally fits this legal requirement. Also, the *Rate of Return Bands/Incentive Regulation* alternative only involves minor changes in the existing rate of return mechanisms, and does not involve substantive changes in other rules.

C. Price Bands

This alternative involves setting both upper and lower limits on prices. Limits on price bands could be set a number of ways, depending on the rationale behind the band. For example, if the rationale for using a price band was to give firms flexibility in price setting, then the band could be set based on a percentage of existing prices. For example, existing prices could be the mid-point of the price band, and prices could be allowed to vary 5% from that mid-point. Obvious issues include how to set the width of the band, and what criteria to use when the entire band needs to be changed.

On the other hand, if the purpose of the price band is specified as a means of preventing monopoly or anticompetitive prices, then the price bands could be based on prices that regulators determined were clearly anticompetitive or monopolistic. For example, the upper limit on the band could be the price that regulators determined would be clearly monopolistic and the lower limit could be the price that regulators concluded was clearly below the costs of providing service, and thus, a predatory price.

Economists have also suggested some criteria for price bands. William Baumol presented an extensive analysis of price floors and ceilings and concluded that compensatory price ceilings and floors (i.e. prices that cover a target revenue requirement) should be based on notions of stand-alone costs.[25] Southwestern Bell Telephone economists Alexander Larson and Calvin Monson recently presented an extensive analysis of price floors and ceilings as applied to telecommunications services.[26]

---

[24] See, e.g., Permian Basin Area Rate Cases, 390 US 747 (1968), reh. den., Bass v. FPC, 392 US 917 and 47 C.F.R. § 65.101 (Oct. 1986).

[25] Baumol, Minimum and Maximum Pricing Principles for Residual Regulation, 5 EASTERN ECON. J. 235 (1979), reprinted in CURRENT ISSUES IN PUBLIC UTILITY ECONOMICS: ESSAYS IN HONOR OF JAMES C. BONBRIGHT 177 (DANIELSON & KAMERSCHEN, eds. 1983)

[26] Larson & Monson, Public Policy Recommendations for the Flexible Pricing of Telecommunications Services, paper presented at the University of Georgia's Seventh Annual Public Utilities conference in Atlanta, GA (Sept. 3, 1987), published in part at 4 TELEMATICS 7 (Dec. 1987).

Price bands are popular regulatory alternatives. Florida, for example, requires AT&T to set its prices within a floor and ceiling, but has generally abandoned earnings constraints.[27] In addition, as of mid-1987, banded rates with both price floors and ceilings for AT&T's intrastate services have been approved or implemented in ten states,[28] and rate ceilings without price floors for AT&T's intrastate services have been implemented in five states.[29] Illinois specifies only a marginal cost-based price floor for competitive MTS, WATS and other switched services.[30] Price bands are allowed for telecommunications services by statute in several states.[31]

1. *Analysis*

A major drawback of rate banding is that it may impede administrative efficiency. How wide should price bands be? How should price bands be changed if circumstances show that price changes outside of the band are necessary? What sanctions should apply if a firm sets its prices outside of the price bands? Should price bands be set for the average price of a service (e.g. price bands for the average prices of MTS and WATS), or should price bands be set for individual rate elements (e.g. the price of additional MTS minutes on weekends)? If the actual rate of return is greater than the allowed rate of return, how should refunds (if any) be apportioned among customers who face banded rates? How should the rate bands be adjusted to prevent over-earning? Obviously, regulatory mechanisms that address these questions run the risk of significantly increasing the administrative burden of regulation.

Rate bands that are set too wide may make consumers worse off to the extent that a regulated firm can increase individual prices and extract monopoly profits from its captive customers. Also, to the extent that a firm possesses market power, a rate band that is set too high creates incentives to increase prices above the costs of providing service. However, this line of analysis assumes that rate of return determined prices better reflect the costs of providing service. In fact, in the telecommunications industry individual service prices are notorious for not reflecting the cost of service. Most commentators argue that toll prices are substantially greater than the cost of providing such services and that local service prices are substantially less than the cost of service.[32] Thus, price banding may bring service prices closer to the costs of service, especially for services, such as toll services, that are offered in competitive markets.

---

[27] Order Granting AT&T Communications of the Southern States, Inc., Forbearance from Earnings Regulation and Waiver of Certain Rules for a Two-Year Trial Period, Docket No. 870347-TI (Aug. 3, 1988). This was the third in a series of "deregulation" hearings in Florida in two years. The Commission held that a marginal cost price floor for MTS and WATS services was administratively burdensome and elected to impose a price floor equal to aggregate access charges plus billing and collection costs.

[28] States Adopt Three Primary Approaches to Easing AT&T's Regulation, 5 STATE TELEPHONE REGULATION REPORT 6, 7-10 (June 18, 1987). These states include Arizona, Michigan, Mississippi, Missouri, New Jersey, New York, Ohio, Oregon, Pennsylvania, and Texas.

[29] Id. These states include Idaho, Louisiana, Montana, North Carolina, and South Carolina.

[30] Id. This is a statutory requirement. See Ill. Ann. Stat. ch. 111 2/3, § 13-507 (Supp. 1988).

[31] See, e.g., Mont. Code Ann. § 69-3-807(2)(e) (1985); Wash. Rev. Code Ann. § 80.36.340 (Supp. 1988)

[32] See, e.g., Temin & Peters, Cross-Subsidization in the Telephone Network, 21 WILLAMETTE L. REV. 199, 210-214 (1985); J. WENDERS, THE ECONOMICS OF TELECOMMUNICATIONS: THEORY AND PRACTICE, 172-177 (1987); and, Bailey, Price and Productivity Change Following Deregulation: The U.S. Experience, 96 ECON. J. 1, 7 (1986).

D. Monopoly Services Regulation

Under this alternative, a rate of return is set for and/or used in conjunction with services where abuse of monopoly power is possible. For example, rate of return regulation could be applied only to services or carriers that could be classified as monopoly or non-competitive, and not to services or firms that are classified as competitive.

This alternative is very popular since many commentators think of regulation and competition as broad alternatives for achieving public policy goals.[33] For example, many recently enacted state statutes permit classification of telecommunications services or firms according to the degree of competition they face, and allow for modification (or elimination) of regulation to reflect such competition.[34] This follows the pattern established at the FCC, which applies rate of return regulation only to dominant carriers (AT&T and local exchange carriers).[35] In a broad sense, the FCC's structural separation rules regarding the provision of enhanced services[36] are another example of this alternative. The rules were based on an FCC finding that the provision of enhanced services is not a natural monopoly, and thus should not be subject to rate of return regulation.[37] Also, in its recent *Streamlined Regulation* docket,[38] the FCC proposed reducing regulatory requirements on services where carriers could demonstrate that they did not possess significant market power.[39]

---

[33] See, e.g., Patrick, Long Distance Carrier Services: Other Modes of Regulation, 119 PUB. UTIL. FORT. 11, 15 (Mar. 5, 1987); and, Policy and Rules Concerning Rates for Competitive Common Carrier Services and Facilities Authorizations Therefor: Further Notice of Proposed Rulemaking, CC Docket 79-252, 84 FCC 2d 445 at para. 11-33 (Jan. 16, 1981).

[34] See, e.g., Colo. Rev. Stat. § 40-15-301 et seq. (Supp. 1988); Fla. Stat. Ann. § 364.337 (Supp. 1988); 1988 Ga. Laws HB-1234; 1988 Idaho Sess. Laws HB-687 [allows deregulation of all but local service]; Ill. Ann. Stat. ch. 111 2/3 §§ 13-502 et seq. (Supp. 1988); Ind. Code Ann. 8-1-2.6-1 et seq. (Supp. 1988); 1988 Ind. Acts SB-239 [deregulates telephone companies with less than 6000 lines]; Iowa Code Ann. § 476.1 (Supp. 1986); Minn. Stat. Ann. sec 237.60 (Supp. 1988); 1988 Miss. Laws HB-1221; Mo. Rev. Stat. §§ 392.010 et seq. (Supp. 1988); Mont. Code Ann. 69-3-807 (1985); N. M. Stat. Ann. §§ 63-9A-1 et seq. (Supp. 1985); N. D. Cent. Code § 49-21-02.1 (Repl. 9B, Supp. 1987); 1988 S.D. Sess. Laws SB-42; Tex. Rev. Civ. Stat., art. 1446c, § 18(a) et seq. (Supp. Vernon 1987); Utah Code Ann. § 54-8b-1 et seq. (Supp. 1985); Va. Code § 56-481.1 (Repl. 1985); Wis. Stat. § 196.195 (Supp. 1987); and, Wash. Rev. Code Ann. secs. 80.36.300-340 (Supp. 1988).

[35] This was the result of several orders than spanned a five year period. See generally, Policy and Rules Concerning Rates for Competitive Common Carrier Services and Facilities Authorizations Therefor: Notice of Proposed Rulemaking, CC Docket 79-252, 77 FCC 2d 308 (1979); First Report and Order, 85 FCC 2d 1 (1980); Further Notice of Proposed Rulemaking, 84 FCC 2d 1 (1980); Second Report and Order, 91 FCC 2d 59 (1982); Second Further Notice of Proposed Rulemaking, 47 Fed. Reg. 17308 (April 22, 1982); Third Report and Order, 48 Fed. Reg. 46791 (Oct. 14, 1983); Fourth Report and Order, 95 FCC 2d 554 (1983); Fourth Further Notice of Proposed Rulemaking, 96 FCC 2d 922 (1984); Fifth Report and Order, 98 FCC 2d 1191 (1984); Sixth Report and Order, 99 FCC 2d 1020 (1985), rev'd. and remanded, MCI Telecommunications Corp. v. FCC, 765 F.2d 1186 (D.C. Cir. 1985).

[36] 47 C.F.R. § 64.702 (Oct. 1986).

[37] Amendment of Sections 64.702 of the Commission's Rules and Regulations (Second Computer Inquiry): Final Decision, CC Docket 85-229, 77 FCC 2d 384 at 433, 440 and note 5, on reconsid., 84 FCC 2d 512 (1981), aff'd. sub nom., Computer and Communications Industry Association v. FCC, 693 F.2d 198 (D.C. Cir. 1982), cert. den., 461 US 938 (1983), Report and Order, 104 FCC 2d 958 (1986) [Third Computer Inquiry].

[38] Decreased Regulation of Certain Basic Telecommunication Services: Notice of Proposed Rulemaking, CC Docket 86-421, 2 FCC Rcd 645 (Jan. 7, 1987). Commissioner Dawson argued that regulation ought to be applied to services that are "core" services. Basically, a "core" service is one that is a substitute for and a basic building

The basic difficulty with this alternative is devising a means of distinguishing between services whose provision is a natural monopoly or non-competitive. By definition, the provision of a service is a natural monopoly if there are economies of scale or scope such that a single provider can meet the demands of the market at lower average costs than could multiple, competing firms.[40] The direct means of making this determination is to estimate the firm's average costs and ascertain whether those average costs decrease with increases in output or when a new product line is added to its production process.[41]

An alternative to calculating average costs is to ask whether the market within which the service is provided could be served by more than one firm, or whether the existing provider possesses market power in that service market. In its *Streamlined Regulation* docket, the FCC proposed a market power test drawn from antitrust law to gauge whether regulated firms possessed market power in service markets being considered for relaxed regulation.[42] Other regulators also employ antitrust principles, such as market definition principles or market power assessment techniques, to determine whether to classify telecommunications services as "competitive".[43]

Most telecommunications regulatory statutes provide explicit criteria to be considered in making a competitive--non-competitive classification. A recent survey of state deregulation legislation, prepared for the Michigan Divestiture Research Board, identified several common criteria found in state statutes to determine whether effective competition exists to warrant relaxed regulation, including: (1) the presence of "reasonable alternatives" to the service in question; (2) the number and size of firms providing competing service; (3) the ease of market entry and availability of competing services; (4) the market power of the existing firms; (5) the importance of the service and service quality to the public; (6) the effect on basic rates; (7) the regulatory burden imposed on telephone companies; and (8)

---

block for all other services, such as local exchange service. See separate Statement of Commissioner Mimi Weyforth Dawson, 4.

[39] The NTIA also recently suggested deregulating services that faced competition or potential competition. NTIA REGULATORY ALTERNATIVES REPORT at 45-68.

[40] *See generally* 2 KAHN, 113-126; PHILLIPS, 38-46; and, Posner, *Natural Monopoly and Its Regulation*, 21 STAN. L. REV. 548 (1969).

[41] This was basically the technique employed in MICHIGAN DIVESTITURE RESEARCH FUND BOARD, 1 TELECOMMUNICATIONS COMPETITION IN MICHIGAN AND REGULATORY ALTERNATIVES (June 1988); and, Petition of the General Counsel for an Evidentiary Proceeding to Determine Market Dominance among Interexchange Telecommunications, Texas Docket 7790, Direct Testimony of Dr. Lee L. Selwyn (May 25, 1988). The extent to which telecommunications firms experience economies of scale and/or scope has been the subject of considerable empirical research. For a survey, *see* Kiss & Lefebvre, *Econometric Models of Telecommunications Firms: A Survey*, 38 REVUE ECONOMIQUE 307 (1987).

[42] Decreased Regulation of Certain Basic Telecommunication Services: Notice of Proposed Rulemaking, CC Docket 86-421, 2 FCC Rcd 645, at pars. 14-16 (Jan. 7, 1987).

[43] *See, e.g.*, Hunt, Market Structure Criteria to Evaluate Lessening Telecommunications Regulation, National Association of Regulatory Utility Commissioners, Committee on Finance and Technology paper (March 20, 1987).

other factors that regulators feel are relevant.[44] Also, the National Regulatory Research Institute (NRRI) has done significant research in this area and has compiled a list of factors and analyses that regulators should consider in evaluating telecommunications competition.[45] Economists have also suggested various methods for analyzing the degree of competition in markets. In telecommunications proceedings, the most popular economic competitive analyses are Baumol, Willig and Panzar's contestable market concepts,[46] Scherer's structure-conduct-performance paradigm,[47] variants of the Lerner Index,[48] and Porter's competitive strategy paradigm.[49]

Prospective deregulation of new services is a variant of the monopoly services regulation theme. Arguably, new services are non-basic, non-core services, so there may be no need to regulate such services. Former Michigan Public Service Commissioner, Matthew McLogan recently proposed abandoning rate of return regulation for new telecommunication services.[50] Under Commissioner McLogan's proposal, firms would be free to set whatever prices they thought appropriate for new services and could earn returns in excess of the allowed rate of return if they shared such excess profits with customers and if they agreed not to pass losses from unsuccessful new ventures onto customers.[51]

---

[44] MICHIGAN DIVESTITURE RESEARCH FUND BOARD, 2 TELECOMMUNICATIONS COMPETITION IN MICHIGAN AND REGULATORY ALTERNATIVES, 45-52 (June 1988).

[45] Horning, et al., Evaluating Competitiveness of Telecommunications Markets: A Guide for Regulators, NRRI report 88-1 (Jan. 1988).

[46] The theory of contestable markets emphasizes the effects of potential market entrants on the pricing and output behavior of firms that might otherwise be considered a monopolist. The theory emphasizes an analysis of entry and exit conditions in a market. See generally, W. J. BAUMOL, R. D. WILLIG & J. C. PANZAR, CONTESTABLE MARKETS AND THE THEORY OF INDUSTRY STRUCTURE (1982); and Willig, Contestable Market Theory and Regulatory Reform, paper presented at University of Texas/Texas Public Utility Commission conference, Current Issues in Telephone Regulation: Dominance and Cost Allocation in Interexchange Markets, Austin, Texas (Oct. 4-7, 1987).

[47] Scherer argues that market structure should be analyzed by examining the structure of the market (e.g. the number of firms, entry barriers), the conduct of firms in the market, and the performance of market in achieving various economic and social objectives. Concepts of "workable competition" are associated with Scherer. See generally, F. W. SCHERER, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE (2d 1980).

[48] The Lerner Index measures the degree to which prices deviate from marginal costs and is based on the notion that prices will be much greater than marginal costs under monopoly conditions than under competitive conditions. The Lerner Index was mathematically related to market share, the elasticity of demand and the elasticity of supply in Landes & Posner, Market Power in Antitrust Cases, 94 HARV. L. REV. 937 (1981). For an application to telecommunications, see Kaserman & Mayo, Deregulation and Market Power Criteria: An Evaluation of State Level Telecommunications Policy, paper presented at University of Texas/Texas Public Utility Commission conference, Current Issues in Telephone Regulation: Dominance and Cost Allocation in Interexchange Markets, Austin, Texas (Oct. 4-7, 1987).

[49] Dr. Porter argues that five forces are relevant to gauging the basic competitiveness in a market: (1) rivalry of existing competitors; (2) threat of new entrants; (3) bargaining power of buyers; (4) threat of substitute services or products; and, (5) the bargaining power of suppliers. See generally, M. PORTER, COMPETITIVE STRATEGY (1980); and, Porter, Competition in the Long Distance Telecommunications Market: An Industry Structure Analysis (Oct. 1987), filed as Appendix A in Policy and Rules Concerning Rates for Dominant Carriers, CC Docket 87-313, Comments of American Telephone and Telegraph Company (Oct. 19, 1987).

[50] McLogan, The Present Function of State Regulation of Local Telephone Service, 119 PUB. UTIL. FORT. 11, 12 (Apr. 16, 1987).

1. *Analysis*

The basic rationale for this alternative is that it focuses regulation on only those services that need to be regulated and uses market forces to assure that rates are "just and reasonable" for competitive services. When such a segregation of competitive and monopoly services is accompanied by mechanisms that prevent cross-subsidization,[52] such a segregation reduces the potential for anticompetitive acts and enhances economic efficiency by driving the prices of monopoly and competitive services closer to their individual economic costs. Also, the segregation of services into monopoly and competitive services increases technical efficiency by enhancing incentives to adopt cost-cutting innovations in competitive services to increase profits.

On the other hand, classification of services as monopoly and competitive may decrease economic efficiency by creating a set of arbitrary allocation rules. One of the major criticisms of traditional regulation is that arbitrary allocations of costs among services creates a pricing system that is unrelated to the actual economic costs of service.[53] If services are segregated into monopoly and competitive classifications, a similar allocation of costs would likely occur to ensure that the monopoly customers were not burdened by the firm's competitive offerings. Such an allocation would likely be arbitrary since economics or accounting offer no unique means of allocating common costs between competitive and non-competitive services.

The effort to segregate monopoly services from competitive services could retard administrative efficiency by resulting in more regulation, not less. Professor Faulhaber argues that service-by-service deregulation results in increased regulation by introducing "the quagmire of separating monopoly from competitive elements (by accounting, separate subsidiaries, et cetera) that the FCC failed to resolve with the Bell system in the years prior to divestiture."[54] This author has argued that the regulatory trend in the telecommunications industry is towards proceedings that focus on competition and antitrust related issues and away from traditional regulation's accounting and finance issues.[55] Ironically, and in contrast to the accepted dogma that "competition is more efficient than regulation," the consequences of the transformation of regulatory proceedings from a revenue requirement focus to an antitrust focus may well be more resources devoted to regulation, and not less.

---

[51] The FCC's net revenue test, as applied to AT&T's new service offerings, is an alternative to complete deregulation of new services. See generally, Guidelines for Dominant Carriers' MTS Rates and Rate Structure Plans: Memorandum Opinion and Order, CC Docket 84-1235, 50 Fed. Reg. 42945 at pars. 26-57 (Oct. 17, 1985). Basically, the net revenue test is a complex set of tests for showing that a new service is not being priced anticompetitively or offered to replace an existing service. If a service passes the net revenue test, it may be offered at competitive rates under relaxed rate of return regulation.

[52] Two mechanisms are commonly employed to prevent cross-subsidization: (1) require that the firm establish a separate subsidiary to provide competitive services; and, (2) require a showing that the competitive services are offered at prices greater than the marginal costs of producing such services.

[53] Traditional rate of return, "cost based" rates fail to reflect the underlying economic costs of service for several reasons. First, revenue requirements are, at best, the historical total costs of service, not a firm's current or anticipated costs. Second, rate base valuations and depreciation calculations do not often reflect the current economic value of an asset, but, at best, the historical book value of assets. In many cases, depreciation rates are chosen to avoid reflecting the true economic depreciation of assets. Third, jurisdictional separations assign arbitrary portions of a firm's total revenue requirements to federal and state jurisdictions. Fourth, fully distributed cost allocation methods assign arbitrary portions of jurisdictionally allocated revenue requirements to individual services. Finally, rates for some telecommunications services, such as local service, are residually priced -- that is, whatever costs are left over after all other costs (revenue requirements) have been assigned to other services are assigned to residually priced services.

[54] Faulhaber, The FCC's Path to Deregulation: Turnpike or Quagmire?, 120 PUB. UTIL. FORT. 22 (Sept. 3, 1987).

Classifying services as monopoly and competitive also suffers from circularity. A service may be monopolistic because regulators do not allow it to be competitive. Market power tests, traditionally applied in antitrust cases,[56] which are a common criteria suggested for classifying services as monopoly or competitive, simply measure the market structure regulators have allowed and, thus, provide no insight into what market structure/regulation regulators ought to apply to a service or market. Also, simply measuring market share or market power typically provides little insight into whether a firm can or will act in an anticompetitive manner or abuse its monopoly power.[57]

The separation of monopoly and competitive markets is often justified because of the danger that the firm will use profits from the monopoly market to cross-subsidize the competitive markets in an anticompetitive manner, or will use the monopoly market as a "bottleneck" to impede competition in the competitive market. Such a justification is questionable for at least three reasons: First, anticompetitive cross subsidization is rarely proved in antitrust cases, so arguably it is not a problem requiring elaborate regulatory redress.[58] Second, imposing regulation to avoid anticompetitive acts assumes that in the absence of regulation the firm would engage in such acts, and, this assumption may be contrary to "real world" telecommunications firms acute awareness of antitrust laws. Finally, antitrust laws were designed to deal with anticompetitive behavior and have done so for almost 100 years; why are these institutions no longer adequate to deal with anticompetitive acts such that a new regulatory institution is now required?

On the other hand, the telecommunications industry presents an economic environment that is radically different than that in a "generic" antitrust case. First, the industry is noted as populated by firms that have a history of having attempted to engage in anticompetitive pricing and cross-subsidization. Second, the relative market shares of firms in the industry are due in large part to 40 years of government protection, and not necessarily due to firms' business acumen. Third, the relative market shares of dominant and non-dominant firms in the telecommunications industry are orders of magnitude apart -- AT&T has between 70 and 80 percent of the market, whereas the next closest competitor has less than 10 percent of the market. Finally, unlike other industries, telecommunications competitors are often placed in a position of having to buy a necessary input from their competitors. For example, interexchange firms must purchase access services from local exchange carriers, and compete with those same carriers in intraLATA markets.

---

[55] Sievers, The Transition to Antitrust Regulation in the Telecommunications Industry: A Comparison of Public Utility Law Regulation and Antitrust Law Regulation, paper presented to Communications Subcommittee of the National Association of Regulatory Utility Commissioners (July 19-20, 1987).

[56] For descriptions of antitrust market power tests, see Landes & Posner, Market Power in Antitrust Cases, 94 HARV. L. REV. 939 (1981); Comment, Landes and Posner on Market Power: Four Responses, 95 HARV. L. REV. 1787 (1982) [price/marginal cost relationships]; Hovenkamp, Antitrust Analysis of Market Power, With Some Thoughts About Regulated Industries, and Flynn, Legal Approach to Market Dominance: Assessing Market Power in Antitrust Cases, Comments on Paper by Herbert Hovenkamp, papers presented at University of Texas/Texas Public Utilities Commission conference, "Current Issues in Telephone Regulation: Dominance and Cost Allocation in Interexchange Markets", Austin, Texas (Oct. 4-7, 1987).

[57] An exception to this is Frentrup, The Effect of Competition and Regulation on AT&T's Intrastate Toll Prices, and of Competition on Bell Operating Company IntraLATA Toll Prices, FCC Common Carrier Bureau, Industry Analysis Division paper (June 22, 1988), where the author presented a model that related AT&T's market share to its MTS prices.

[58] See, e.g., Koller, The Myth of Predatory Pricing -- An Empirical Study, 4 ANTITRUST L. & ECON. REV. 105 (1971); and Matsushita Electric Industrial Co. Ltd. v. Zenith Radio Corp., 475 U.S. 574 (1986).

Another regulatory concern about the segregation of competitive and monopoly services is the impact on local service rates. The principle fear is that the loss of local service subsidies from competitive services will drive up local rates. There has been little empirical research done in this area,[59] but most of what has been done indicates that the segregation of markets and the introduction of competition in long distance markets has not caused dramatic increases in local service rates. For example, in its study of the pricing impact of the introduction of interexchange competition, the FCC concluded that

> [d]uring the three years since the AT&T divestiture (which occurred on January 1, 1984), the overall price of telephone services purchased by residential customers has risen at a slower rate than during the three years before divestiture. The average annual rate of increase in the overall cost of telephone service has been 5.5% since divestiture. This is almost identical to the average rate of increase for the six years prior to divestiture. Important differences have developed, however, between changes in local and long distance rates. During the 1970's and early 1980's, local and long distance rates tended to follow the same patterns. In 1981, for example, both local rates and interstate long distance prices increased by record amounts and the percentage increases were similar. By 1984, however, the pattern of similar movements had been broken. During 1984, 1985, and 1986, the price of interstate long distance service declined. During those same years, the price of local service continued to increase, albeit at a declining rate.[60]

III. MAJOR REGULATORY CHANGES

This section considers five alternatives that abandon or radically alter basic revenue requirements structure:

A. *Revenue Regulation Only.* Revenue requirements are not used to set any individual service prices, but are used only to set an overall revenue ceiling.

---

[59] See, e.g., Noll & Riley, The Laboratory of the States: Local Service Prices Before and After Divestiture, paper presented at Conference on Utility Regulation, Stanford University Center for Economic Policy Research (April 21-23, 1988).

[60] Lande & Wynns, Primer and Sourcebook on Telephone Price Indexes and Rate Levels, FCC Common Carrier Bureau, Industry Analysis Division report at par. 5 (April 1987). For an update to these statistics, see Lande, Local Rates Update (Sept. 1987 & Dec. 1987).

B. *Price Regulation Only.* Only the price of individual services are regulated, not overall earnings or revenues. Price regulation could take the form of price caps, or price standards.

C. *Sanctions/Deregulation.* A set of sanctions are structured to prevent regulated firms from abusing their market power, but otherwise there is no direct regulation of prices or revenues.

D. *Public Ownership.* Monopoly services, such as local telephone service, are owned and operated by local or state government, similar to the state ownership of telephone companies in Europe and local government ownership of electric, gas and water utilities in the United States.

E. *Franchise Bidding.* Firms would have to periodically bid and pay for the right to provide monopoly service in a specified area, and, if awarded the franchise, would be free from price or earnings regulation. The proceeds from the bid could be used to advance social goals and/or offset excessive prices.

A. Revenue Regulation Only

Rate of return regulation is composed of two broad, related forms of regulation: total revenue regulation, and service price regulation. Each class of regulation is arguably directed at different regulatory goals. Broadly, revenue regulation seeks to limit the firm's ability to extract monopoly profits from its customers, while ensuring that the firm's revenues are sufficient to cover its total costs and pay investors "just compensation" for property devoted to public use. Price regulation, on the other hand, is typically concerned with allocating a firm's revenue requirements among individual services.[1] Many of the economic criticisms of rate of return regulation, such as the creation of incentives for uneconomic bypass and inefficient cross-subsidization, are due to a failure of the price regulation component of regulation, and not necessarily revenue regulation. Thus, a good deal of economic distortion could be eliminated by eliminating service price regulation, while retaining overall revenue regulation.

The "marketbasket" approach to revenue regulation is a variant of this alternative.[2] Under the "marketbasket" alternative, only the firm's total profits are regulated to prevent the firm from earning monopoly profits. The firm's profitability under the "marketbasket" approach is measured by a variety of indicators in much the same manner as a firm's credit rating is established by reference to financial ratios and industry norms. If there is evidence that the firm earns monopoly profits, then regulators could order a refund of excess earnings to ratepayers. The rationale for the "marketbasket" approach is that it is consistent with how other firms are evaluated by creditors, and since no explicit revenue requirements constraint is involved, it avoids the over-capitalization incentives of traditional revenue requirements regulation.

---

[1] In general rate proceedings, this allocation is often referred to as "rate spread" or "rate design." See the courts' discussions in Midwest Gas Users Association v. State Corporation Commission, 623 P. 2d 924 (Kan. App. 1981) and Texas Alarm and Signal Association v. Public Utility Commission of Texas, 603 S.W. 2d 766 (Tex. 1980) for the distinction between rate of return regulation and rate design.

[2] See generally, Comprehensive Review of Rate of Return Regulation of the Telecommunications Industry: Notice and Request for Comments, National Telecommunications and Information Administration docket no. 61091-6191, 51 Fed. Reg. 36,837 (Oct. 16, 1986). Also see, Harlow, Efficient Market Perspectives on Utility Rate of Return Adequacy, 113 PUB. UTIL. FORT. 38 (Mar. 29, 1984) who advocates use of a Q-Ratio as a marketbasket indicator.

1. *Analysis*

An important advantage of the *Revenue Regulation Only* alternative is that it would increase administrative efficiency by eliminating price regulation, a major, administratively burdensome component of traditional regulation. This elimination of price regulation would also decrease investor uncertainty by decreasing the regulatory risk associated with the price setting process. However, as Judge Posner points out, the deficiency of using only aggregate rates of return as measures of monopoly power is that monopolized industries often exhibit normal or below normal aggregate returns, but individual prices in those industries are often far in excess of the underlying economic costs.[3] Thus, the *Revenue Regulation Only* alternative might do little to control anticompetitive acts, such as excessive or below cost prices.

The *Revenue Regulation Only* alternative could also conflict with regulatory statutes that require regulators to set "just and reasonable" rates. For example, in *FPC v. Texaco, Inc.*,[4] the Supreme Court held that the Federal Power Commission could not rely exclusively on market forces to set just and reasonable rates for natural gas since that conflicted with the legislative assumption that natural gas markets are not competitive markets and some form of rate regulation is necessary.[5] Thus, some form of rate regulation seems legally necessary when a statute requires regulators to set "just and reasonable" rates.

B. Price Regulation Only

In contrast to only revenue regulation, this alternative employs strictly price regulation. The rationale for such an approach is that: (1) regulatory statutes typically require "just and reasonable" rates and services, but are silent as to total revenues,[6] so arguably revenue regulation is excess regulatory baggage; and, (2) economic distortions are introduced in the pricing of individual services, not necessarily from revenue regulation, so regulatory rules ought to focus on those pricing rules. Broadly, price regulation alternatives can be placed into two categories:

1. *Price Caps.* Existing prices are presumed to be "just and reasonable" and initially used to set price caps, which are basically upper limits on the average price of a collection of services. Firms are allowed to change any prices with minimal (or without) regulatory approval so long as the price cap is not violated. Price caps are automatically moved upwards according to inflation, as measured by

---

[3] R. POSNER, ANTITRUST LAW: AN ECONOMIC PERSPECTIVE, 15 (1976).

[4] 417 U.S. 380, 94 S.Ct. 2315, 41 L.Ed. 2d 141 (1974).

[5] Id. 94 S.Ct. at 2326-2328. Also see, Farmers Union Central Exchange, Inc. et al. v. FERC, 734 F. 2d 1486 (D.C. Cir. 1984); and generally, Note, Storming the AT&T Fortress: Can the FCC Deregulate Competitive Common Carrier Services?, 32 FED. COMM. L. J. 205 (1980).

[6] For example, 47 U.S.C. § 201(b) of the Communications Act, which governs the FCC's rate regulation refers to individual prices, not total revenues. "All charges, practices, classifications, and regulations for and in connection with such communication service shall be just and reasonable, and any charge, practice, classification, or regulation that is unjust and unreasonable is declared to be unlawful."

a suitable price index and may also be adjusted according to changes in industry productivity or exogenous changes in total costs.

2. *Price Standards.* Prices are set or changed according to some predefined standard, such Ramsey prices, incremental costs, Pareto standards, Kaldor-Hicks standards, fairness standards, or other economic standards.

1. *Price Caps*

Recently, the FCC formally proposed price cap regulation as an alternative to traditional rate of return regulation for dominant telecommunications carriers,[7] which was has resulted in a number papers analyzing price cap regulation.[8] Price cap regulation uses existing rates to calculate service price caps, which are basically constraints on the average price of service for a collection of services (called the price cap "basket").[9] Any changes in individual service prices are allowed so long as the price cap for the basket is not violated. The price cap is automatically adjusted according to inflation, industry productivity and exogenous cost changes.

The price cap proposal currently under review by the FCC uses the percentage point change in the GNP producer price index (GNP-PI) to measure inflation.[10] In its proposal, the FCC examined a great deal of empirical evidence and concluded that the average annual industry productivity gain for both AT&T and local exchange carriers is 2.5 percent.[11] The FCC also proposed an additional annual adjustment to the price cap of 0.5 percent, which it referred to as the *consumer productivity dividend.*[12] The *consumer productivity dividend* reflects the FCC's belief that carriers will be more efficient under price cap regulation, and the dividend is intended to share those productivity gains with consumers. The FCC also proposed to adjust the price cap by changes in revenue requirements that have already been ordered by the FCC.[13] While the FCC's proposal is the subject of an ongoing comment cycle and considerable academic debate, its proposed adjustment can be algebraically represented as

---

[7] Policy and Rules Concerning Rates for Dominant Carriers: Notice of Proposed Rule Making, CC Docket 87-313, 2 FCC Rcd 5208 (1987); ____: Further Notice of Proposed Rulemaking, ___ FCC Rcd. ____ (May 23, 1988).

[8] The most influential price caps papers include: Patrick, Long Distance Carrier Services: Other Modes of Regulation, 119 PUB. UTIL. FORT. 11, 13-15 (Mar. 5, 1987) [proposed price caps]; Haring & Kwerel, Competition in the Post-Equal Access Market, 2 FCC Rcd 1488 (Mar. 11, 1988) [economic analysis of price caps]; Levitz, Loosening the Ties that Bind: Regulating the Interexchange Services Market for the 1990's, 2 FCC Rcd. 1495 (Mar. 12, 1988) [legal analysis of price caps]; Eagan & Taylor, The Economics of Ceiling Price Regulation, paper presented at Rutgers University Advanced Workshop in Public Utility Economics and Regulation, Western Conference (July 6-8, 1988) [Bellcore theoretical analysis of price caps]; and, Bellcore (Taylor), The Impact of Federal Price Cap Regulation on Interstate Toll Customers, Bell Communications Research paper (March 1988) [Bellcore empirical analysis of price caps].

[9] Technically, the price cap in any given year is calculated as the average price using the current year's prices and the base year quantities. Some price cap proposals call for the base year quantities to be fixed for some "contract period", generally three to five years, whereas other allow the change the base year quantities annually.

[10] Policy and Rules Concerning Rates for Dominant Carriers: Further Notice of Proposed Rulemaking, CC Docket 87-313, ___ FCC Rcd ___, at pars. 346-355 (May 23, 1988). [hereafter cited as FURTHER NOTICE]

[11] FURTHER NOTICE at pars. 357-396.

[12] FURTHER NOTICE at pars. 386-389.

$$P_t = P_{t-1}\,(1 + \text{GNP-PI} - 0.025 - 0.005 + Y)$$

where

P_t is the current year's price cap, and P_{t-1} is last year's price cap;

GNP-PI is the measure of inflation;

0.025 is the FCC conclusion about productivity;

0.005 is the *consumer productivity dividend*; and,

Y is the exogenous revenue requirement changes expressed as a percentage of base year revenue requirements.

The FCC also proposed that there be two price cap baskets. For the local exchange carriers, the FCC proposed a price cap basket for switched access services and one for special access services. For AT&T the FCC proposed that there be a price cap basket for all of its services that use the switched network and a separate basket for its private line services.[14] In order to protect consumers from dramatic increases in individual service prices, and to protect competitors from predation, the FCC proposed that the prices of all rate elements be restricted within a five percent price band. In addition, the FCC proposed to allow price decreases of greater than five percent, but would not count decreases of greater than five percent in determining compliance with the current price cap.[15]

The commonly cited rationales for price cap regulation include:

1. *Administrative Simplicity.* Price cap regulation is simpler to administer than traditional rate of return regulation.

2. *Enhanced Efficiency Incentives.* Because price cap regulation does not tie revenues and profits to the cost of service, it creates economic incentives for firms to reduce costs and increase network usage. Some economists have argued that a price cap adjustment mechanism will iterate to economically efficient prices and is especially suited to situations where regulators have limited information about regulated firms' costs.[16]

---

[13] FURTHER NOTICE at para. 397-423. The exogenous costs that the FCC considered included: (1) access charge changes previously ordered by the Commission; (2) changes in tax law that affect revenue requirements; (3) changes in jurisdictional separations previously ordered by the Commission; (4) changes in international accounting rates; and, (5) changes in depreciation rates and amortization amounts previously ordered by the Commission.

[14] FURTHER NOTICE at para. 278-284.

[15] FURTHER NOTICE at para. 284-291.

[16] See, e.g., Vogelsang & Finsinger, A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms, 10 BELL J. ECON. & MGT. SCI. 157 (1979); Brennan, Capping Average Prices of Regulated Multiproduct Firms (Oct. 1987); and, Vogelsang, Price Cap Regulation of Telecommunications Services: A Long-Run Approach, Rand Corporation report N-2704-MF (Feb. 1988). However, also see Bradley & Price, Regulation Through an Average Revenue Constraint, University of Leichester working paper (March 1987) which reaches a contrary conclusion.

3. *Guaranteed Average Real Price Decreases.* Unlike traditional rate of return regulation, where prices are based on a measure of the costs of service, price cap regulation guarantees consumers average real price decreases at least as great as the price changes realized under rate of return regulation.[17]

4. *Elimination of Cross-Subsidization Incentives.* Since rates set under price cap regulation are not based on regulatory measures of service costs, price cap regulation eliminates incentives to engage in cross-subsidization by shifting costs between service categories, so long as competitive services are not included in the same basket as non-competitive services.

Britain's Office of Telecommunications (OT) uses a price cap mechanism to regulate British Telecommunications.[18] Under its RPI-3 formula, changes in the average price of access, maintenance and local telephone services are limited to three percentage points less than the annual changes in the retail price index. The plan was adopted to avoid the costs of "American style regulation" -- drawn out, legalistic proceedings and "cost plus" incentives created by rate of return regulation.

Variants of price cap regulation can be found at the state level, as well. For example, in Oregon, Pacific Northwest Bell recently proposed to freeze local rates for two years, peg local rates to an Oregon price index, and share excess revenues with consumers. Southwestern Bell Telephone recently proposed a price index plan for Missouri that indexes price changes in local service, Centrex and PBX rates to the Consumer Price Index less three percent.[19]

2. *Price Standards*

In contrast to price caps and price bands, which regulate prices by specifying a range of "just and reasonable" prices, price standards seek to regulate prices by specifying "the" price that meets a predefined standard.[20] Traditional rate of return regulation basically employs a price standard, since it consists of rules that define how prices are to be calculated.

There are many price standards that could be employed as alternatives to rate of return regulation. For example, Ramsey prices are popular among economists as a means of pricing services such that total revenues are sufficient to cover total costs.[21] Incremental costs, which are calculated by many telephone companies and used as price floors, could also be used as a standard for setting prices outside of a revenue requirements formula.

---

[17] See generally, Taylor, The Impact of Federal Price Cap Regulation on Interstate Toll Customers, Bell Communications Research, Inc. paper (March 1988).

[18] See generally, Stelzer, Regulating Telecommunications in Britain: A New Alternative to the US Approach, 3 TELEMATICS 7 (Sept. 1986); Wheatley, British Experience with Price Regulation, paper presented at Bell Communications Research Conference, Washington D.C. (Aug. 24, 1987); and, Wheatley, Implementation of Price Caps, paper presented at 7th International Conference of the International Telecommunications Society, Cambridge, MA (June 29 - July 1, 1988).

[19] Bocs to Try Deregulation Anew in Maryland, Vermont; More Pending, 6 STATE TELEPHONE REGULATION REPORT 1, 3-4 (Aug. 11, 1988).

[20] For a discussion of various cost-price standards for telecommunications services, see generally Guidelines for Dominant Carrier's MTS Rates and Rate Structure Plans: Memorandum Opinion and Order, CC Docket 84-1235, 50 Fed. Reg. 42945 (Oct. 17, 1985) and, in an antitrust context, see MCI Communications v. AT&T, 708 F.2d 1081, 1111-1131 (7th Cir. 1983), cert. den., 464 U.S. 891 (1983), appeal after remand, 748 F.2d 799 (1984).

Welfare economics standards could be also be employed as alternative standards for guiding price setting. For example, prices could be frozen at present levels and the firm could be required to show that any changes in its prices do not make any one worse off (a Pareto standard) or that the price change results in a net welfare gain to society (Kaldor-Hicks standard).

Economists have proposed a number of price standards.[22] Cross suggested a price maximum equal to some minimum level plus a proportion of the average total costs associated with the service.[23] Sudit[24] and Baumol[25] suggested a system of price standards that tracks changes in input prices to reflect changes in factor productivity.

Loeb and Magat suggested letting the utility set its prices and providing the utility with a subsidy equal to the consumer surplus generated by the chosen price in order to provide incentives for the utility to choose the price structure that maximizes consumer welfare.[26] In order to minimize the sums ultimately recovered by utilities, Loeb and Magat suggested simultaneously imposing a lump sum tax on utilities to strip away inequitable wealth transfers to utilities.

Vogelsang and Finsinger suggested setting prices such that zero economic profits would be earned at last year's costs.[27] They asserted that over time, the welfare maximizing price structure would emerge from such a process.

Laffont and Tirole suggested a very complex price standard that is based on the out-of-pocket costs associated with a service plus a transfer payment that depends on marginal costs and total costs reported by the regulated firm.[28] Generally, the transfer payment would be the fixed charge part of a tariff and would decrease with larger announced total costs and would increase with decreases in marginal costs and output.

---

[21] For an extensive discussion of Ramsey pricing in telecommunications, see generally, BROWN & D. SIBLEY, THEORY OF PUBLIC UTILITY PRICING (1987).

[22] Three good surveys of alternatives to rate of return regulation are Brown, Alternatives to Rate of Return Regulation, paper presented to the National Association of Regulatory Utility Commissioners, July 1987; Joskow & Schmalensee, Incentive Regulation for Electric Utilities, 4 YALE J. on REG. 1 (1986); and, Crew & Kleindorfer, Governance Structures for Natural Monopoly: A Comparative Institutional Assessment, 14 J. BEHAVIORAL ECON. 117 (1985).

[23] Cross, Incentive Pricing and Utility Regulation, 89 Q. J. ECON. 237 (1970).

[24] Sudit, Automatic Rate Adjustments Based on Total Factor Productivity Performance in Public Utility Regulation, in PROBLEMS IN PUBLIC UTILITY ECONOMICS AND REGULATION 55 (Crew. ed. 1979).

[25] Baumol, Productivity Incentive Clauses and Rate Adjustment for Inflation, 115 PUB. UTIL. FORT. 11 (July 22, 1982).

[26] Loeb & Magat, A Decentralized Method for Utility Regulation, 22 J. L. & ECON. 399 (1979). Consumer surplus measures aggregate willingness to pay for a service above what was actually paid for a service. Intuitively, it measures the total market value of a service. It is measured as the area under a demand curve at a given price/quantity pair, minus total consumer expenditures for the service.

[27] Vogelsang & Finsinger, A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms, 10 BELL J. ECON. & MGT. 157 (1979).

[28] Laffont & Tirole, Using Cost Observation to Regulate Firms, 94 J. POL. ECON. 614 (1986).

Schleifer suggested using the prices of other comparable, but competitive firms as a "yardstick" for regulated firms' prices.[29] Since exactly comparable firms generally cannot be found for public utilities, Schleifer suggested regressing the prices of competitive firms against various cost variables in order to construct a formula for emulating the prices the firm would have charged if it were in a competitive environment.

3. *Analysis*

The major weakness of the *Price Standards* alternatives is that, a priori, they appear to offer no administrative improvement over existing price standards. While price standards, such as stand alone-costs or incremental costs, seem attractive, no one has ever done such cost studies independent of revenue requirement calculations, nor has anyone compared the administrative costs of the existing price standards with such cost based price standards. Such cost-based price standards typically require some sort of rate of return calculation to estimate either the cost of capital embedded in a service or the embedded opportunity costs of a service. Thus, such cost-based price standards probably would result in administrative efficiency losses since they would entail a greater regulatory burden; they require a rate of return calculation plus an additional set of calculations. Also, regulatory proceedings would be more complicated under such price standards as regulators and regulated firms would be required to struggle with the methodology and calculations involved in addition to traditional conflicts arising out of the rate of return calculations.

Some price standards could be simpler than existing rules. For example, regulators could adopt welfare criteria, such as the Pareto or Kaldor-Hicks criteria, and require as a prerequisite of a price change that the firm show that no customer class is made worse off (Pareto standard) or that society realizes a net gain from the price change (Kaldor-Hicks standard).[30] However, such standards are notoriously difficult to meet and thus stifle rate changes and would impair economic efficiency. Such standards are also subject to a good deal of interpretation and manipulation, so they would increase regulatory risk and investor uncertainty.

Price Caps, on the other hand, are simple rules that would result in substantial administrative efficiency gains since they would largely eliminate many of the revenue requirement calculations. The average price ceiling could be automatically determined as the last year's price cap plus the percentage change in a widely accepted measures of inflation and productivity. Prices that satisfy the price cap would be presumed to be "just and reasonable" so they could be implemented without the attendant costs associated with a rate change under the existing rules.

Empirical evidence indicates that consumers would be better off under price cap regulation than under traditional rate of return regulation. One study estimated that interstate access charges would have been about $995 million lower if price cap regulation was implemented at divestiture and applied to both the local exchange carriers and AT&T.[31] It should be noted that this figure excludes any efficiency gains that might be realized by price cap regulation.

---

[29] Schleifer, A Theory of Yardstick Competition, 16 RAND J. ECON. 319 (1985).

[30] For example, the argument that $2 Subscriber Line Charges resulted in average net decreases in customers' monthly bills satisfies a Kaldor-Hicks welfare standard. See Sievers, But Is It Fair? Economic Fairness Standards Applied to Federal Telecommunications Access Charges, paper presented at Rutgers University Advanced Workshop in Regulation and Public Utility Management (March 23, 1987).

The FCC's price caps proposal has been the subject of extensive criticism in Congress. The National Association of Regulatory Utility Commissioners and others complain that service quality might suffer under the FCC's price caps proposal.[32] New York Commissioner Schwartz argues that price cap regulation fails to provide consumers with their fair share of efficiency gains and profits. She proposes a revenue sharing mechanism.[33] US Sprint has argued that including competitive and non-competitive services in the same price cap basket creates incentives for costless, anticompetitive cross-subsidization -- by financing the below-cost offerings with modest increases in non-competitive service prices, a firm could price its competitive services below cost to drive out market rivals without violating the basket's price cap and without suffering any revenue losses.

Professors Cabral and Riordan argue that while price cap regulation may create cost reduction incentives for dominant carriers (AT&T), it may not enhance the cost reduction incentives of non-dominant, competing carriers.[34]

Limiting regulation to only price regulation may not be politically feasible since it is contrary to traditional concerns about excess profits. For example, in Congressional hearings to investigate the FCC's price caps proposal, the Maryland Peoples Counsel, a state consumer representative, focused on AT&T's profits since divestiture, not its prices.[35] Similarly, the Consumer Federation of America's principle complaints about relaxation of regulation and price caps have focused on the excess profits earned by AT&T and the Bell Operating Companies, not the prices charged to consumers or the general affordability of basic telephone service.[36]

C. Sanctions/Deregulation

In traditional rate of return regulation, and with the alternatives described above, economic regulation takes the form of affirmative government supervision; regulators are actively involved in setting revenues and prices. This active supervision is in contrast to *unregulated* markets where prices and revenues are not supervised by a government entity. However, even *unregulated* firms are subject to some government regulation in the form of antitrust sanctions and penalties that might be applied if *unregulated* firms abuse their market power or engage in unfair competitive practices. Thus, an important difference between *regulated* and *unregulated* firms is that the latter are regulated with sanctions, whereas *regulated* firms are subject to active government supervision.

---

[31] Taylor, The Impact of Federal Price Cap Regulation on Interstate Toll Customers, Bell Communications Research, Inc. paper (March 1988). The study is based on the USTA's price cap proposal.

[32] Telephone Service Quality Under Price Caps, discussion paper for Staff Subcommittee on Communications at NARUC Summer meeting in San Diego, CA (July 1988).

[33] Schwartz, Price Caps with a Consumer Benefit Mechanism: A Workable Approach for AT&T, 6 TELEMATICS 1 (June 1988).

[34] Cabral & Riordan, Incentives for Cost Reduction Under Price Cap Regulation, paper presented at Conference on Utility Regulation, Stanford University Center for Economic Policy Research (April 21-23, 1988).

[35] 6 REPORT ON AT&T 7 (Aug. 8, 1988).

[36] CONSUMER FEDERATION OF AMERICA, DIVESTITURE PLUS FOUR: TAKE THE MONEY AND RUN (Dec. 1987).

Under this alternative, rates set by the regulated firm would be presumed to "just and reasonable." This burden of proof is in contrast to traditional rate proceedings where the regulated firm bears the burden of proving that its rates or rate changes are "just and reasonable." In order to insure that the firm does not abuse its monopoly power, antitrust-like sanctions could be imposed by utility regulators on firms that set anticompetitive or monopoly rates.

For example, a regulated firm sets its rates according to its business needs and objectives and files those rates with the Commission. Those rates would automatically become effective when filed or shortly thereafter. In order to ensure that regulated firms do not set monopolistic rates or engage in anticompetitive activities, such as price discrimination, affected parties, such as competitors and consumer groups, could be given the right to bring a quasi-antitrust action against the firm. If an affected party could show that the regulated firm's rates were set at monopolistic or anticompetitive levels, then that party would be entitled to recover its costs of bringing the suit plus its damages. To the extent necessary, the regulator could craft additional remedies, such as price roll-backs, or fines. The same sorts of standards of proof used in antitrust cases could be used by the administrative body in judging whether a firm engaged in anticompetitive acts or abused its monopoly power.

Most states' public utility statutes provide for fines and criminal sanctions for violation of utility law, so this alternative is simply an extension of existing institutions. Under the Communications Act, the FCC is empowered to hear a complaint for damages resulting from any violation of the Communications Act by a common carrier.[37] As another example, Kansas' public utility law statutes provide for treble damages plus costs and attorneys' fees for violations of its public utility laws.[38] The streamlined tariff provisions for competitive services, described above, are another example of this alternative. The rates filed under such tariffs are generally presumed to be "just and reasonable" unless someone comes forward and shows that they are improper.

States like Nebraska and Virginia that have substantially "deregulated" telecommunication services effectively apply this alternative.[39] It is important to note that even though rates and revenues are not regulated, these states continue to monitor rates and revenues with the possibility of re-regulating firms.[40] In a sense, the threat of re-regulation is a sanction.

---

[37] 47 U.S.C. §§ 206-207.

[38] See e.g. Kan. Stat. Ann. § 66-176 (1985). This provision of Kansas law has rarely been applied. See e.g., Kearney v. Kansas Public Service Company, 665 P.2d 757, 769-770 (Kan. 1983) [city gas company held not liable under this provision]; and, Western Kansas Express, Inc. v. Dugan Truck Line, Inc., 720 P.2d 1132 (Kan. App. 1986) [suit between competing trucking firms]. Also see generally, Furney, Recovery of Attorney Fees in Kansas, 18 W. L. J. 535, 562 (1979).

[39] In contrast, some states have gone from a basically unregulated status to regulation. For example, Texas' state public utility law statutes were enacted in 1975. Prior to that time, telecommunications services were largely unregulated by the state, and only municipalities regulated telephone rates in Texas. For a background and history of Texas' utility laws, see Hopper, A Legislative History of the Texas Utility Regulatory Act of 1975, 28 BAYLOR L. REV. 777 (1976); and, Newcomb, Some Aspects of Regulation of Public Utilities Operating in Texas, 5 BAYLOR L. REV. 335 (1953).

1. *Analysis*

Sanctions could result in a decrease in administrative efficiency if they turn into antitrust trials, which are notorious for being extremely long proceedings involving extraordinarily burdensome discovery and compliance costs.[41] To the extent that such sanctions are effective, they could reduce the potential for anticompetitive acts since they directly address such acts in contrast to the oblique approach of revenue and price regulation. Because these alternatives would rely more heavily on market forces to set prices, they would result in greater incentives to adopt price cutting innovations and would result in enhancements to economic efficiency.

For the FCC, sanctions are less feasible than the other alternatives since the FCC is unlikely to adopt an entirely different form of regulation. They would require the FCC to shift from a industry-wide rulemaking body to an adjudicatory body hearing antitrust complaints. A service-by-service deregulation is more likely at the FCC than wholesale deregulation. However, while "deregulation" of local or basic telephone service is unlikely, Nebraska and Virginia have shown that it is possible to substantially "deregulate" major portions of the telecommunications industry.

D. **Public Ownership**

In contrast to the above alternatives, which seek methods of controlling the prices and profits of private firms, this alternative would eliminate monopoly incentives by public ownership of telephone companies. In the electric utility industry, it is common for local distribution facilities to be owned by local government. Also, the federal government operates most of the hydrogeneration facilities in the United States and has a substantial investment in transmission facilities. Most of literature in this area indicates that electric rates in publicly provided systems are substantially lower than the comparable rates of privately owned systems, and there are some indications that public systems are more efficient than private electric systems.[42]

While the public ownership and operation of a telephone system is rare in the United States,[43] it is common in other countries. The general perception of the experience in other countries tends to be that public ownership is generally inferior to private industry, and consequently, there is a move to "privatize" public utilities in some European countries. However, in an empirical investigation of European telecommunications productivity, Professor James Foreman-Peck found that the difference in ownership is not the principle reason for productivity differences between U.S. and European telecommunications firms. But, he also found that there were increases in British total factor productivity after liberalization and privitization.[44]

---

[40] See, e.g., Sedgley, The Monitoring Function: Telecommunications Regulation for the Future, paper presented at 7th International Conference of the International Telecommunications Society, Cambridge, MA (June 29, 1988).

[41] Easterbrook reported that prior to 1981, AT&T estimated that it spent about $100 million annually to defend itself against predation charges. Estimates of the litigation costs of predation cases range from $3 to $30 million. Easterbrook, Predatory Strategies and Counterstrategies, 48 U. CHI. L. REV. 263, 334-335 (1981).

[42] See, e.g., Fare, Grosskopf & Logan, The Relative Performance of Publicly-Owned and Privately-Owned Electric Utilities, 26 J. PUB. ECON. 89 (1985); Meyer, Publicly Owned Versus Privately Owned Utilities: A Policy Choice, 57 REV. ECON. & STAT. 391 (1975); Neuberg, Two Issues in the Municipal Ownership of Electric Power Distribution Systems, 8 BELL J. ECON. & MGT. SCI. 303 (1977); Shenk, Basic Electric Service Costs Less in Public Power Cities, 44 PUBLIC POWER 28 (Jan.-Feb. 1986); Bruneel, New Retail Rate Study Confirms Benefits of Municipal Ownership, 44 PUBLIC POWER 18 (Jan.-Feb. 1986); and, Pescatrice & Trapani, The Performance and Objectives of Public and Private Utilities Operating in the United States, 13 J. PUB. ECON. 259 (1980).

Public ownership of telephone companies is not likely to occur in the United States, since it would require either condemnation of telephone companies or building a telephone network. Also, public ownership would require establishment of new public agencies that are intimately familiar with the technical operations of telephone networks.

E. **Franchise Bidding**

Franchise bidding has been suggested by a number of prominent economists as an alternative means of controlling monopoly profits.[45] Basically, a firm would be required to periodically bid and pay for the exclusive right to serve a particular community. The firm would be subject to very little, or only local rate regulation. Franchise bidding has been applied to the cable television industry,[46] and conceivably could be applied to local telephone service.

As with the public ownership alternative, this alternative does not seem feasible given that franchise rights have already been assigned to various telephone companies. Also, since local telephone rates are widely believed to be less than the costs of providing such service, a franchise bidding scheme might result in substantial increases in local service, which further inhibits its feasibility.

---

[43] During World War I, telephone systems were virtually nationalized and operated by the Postmaster General. An excellent history of federal communications regulation is in Loeb, The Communications Act Policy Toward Competition: A Failure to Communicate, 1978 DUKE L. J. 1 (1978).

[44] Foreman-Peck, Industry Structure in European Telecommunications: The Impact of Liberalization, paper presented at 7th International Conference of the International Telecommunications Society, Cambridge, MA (June 28 - July 1, 1988).

[45] Demsetz, Why Regulate Utilities?, 11 J. L. & ECON. 55 (1968); and, Posner, The Appropriate Scope of Regulation in the Cable Television Industry, 3 BELL J. ECON. & MGT. SCI. 98 (1972). Also see, Goldberg, Regulation and Administered Contracts, 7 BELL J. ECON. & MGT. SCI. 426 (1976) who analyzes the franchise bidding proposal and argues that regulation is akin to an administered contract.

[46] See the case study of Oakland, California described in O. E. WILLIAMSON, THE ECONOMIC INSTITUTIONS OF CAPITALISM, 326-364 (1985).

TABLE 1

EVALUATION CRITERIA

CRITERIA	DESCRIPTION
1. **Protect Customers**	An alternative should leave customers at least as well off as they are under existing institutions, and possibly better off. Protection of universal service and promotion of service to selected customer classes that might be adversely affected by an alternative are included under this criteria.
2. **Protect Investors**	An alternative should leave investors at least as well off as they are under existing institutions, and possibly better off.
3. **Promote Fairness**	An alternative should be fair to customers, investors, competitors and other affected parties.
4. **Promote Competition**	An alternative should promote competition and reduce incentives for dominant firms to engage in anticompetitive acts, such as monopoly pricing (i.e. excessive rates or returns) or predatory pricing (rates designed to injure competitors).
5. **Enhance Economic Efficiency**	An alternative should enhance economic efficiency by driving prices closer to the economic costs of providing service.
6. **Enhance Administrative Efficiency**	An alternative should enhance the administrative efficiency of existing regulation, or at least not add to the complexity and burdens of existing regulation.
7. **Promote Innovation Maintain Service Quality**	An alternative should increase incentives for firms to adopt innovative means of providing service, while maintaining existing high levels of service quality.
8. **Implementation**	An alternative should be politically and legally feasible.

TABLE 2

ALTERNATIVES TO RATE OF RETURN REGULATION FOR THE TELECOMMUNICATIONS INDUSTRY

ALTERNATIVE	
I. MINOR REGULATORY CHANGES	Alternatives that retain traditional revenue requirements structure, but involve minor modifications to the calculation and/or use of revenue requirements.
A. **Negotiated/Abbreviated Rate Proceedings**	Elements of revenue requirements or rates are determined outside of a formal regulatory environment.
1. Social Contract	Regulators and utilities negotiate with each other and arrive at mutually agreeable rates, revenues and conditions of service for a broad range of services. Often involves a price freeze on local service in exchange for pricing freedom on other services.
2. Fast Track Proceedings	Utility is allowed to submit tariff changes for selected services on a streamlined or expedited approval. Often takes the form of streamlined regulation of competitive services and mechanisms similar to automatic fuel adjustment clauses in the electric industry.
B. **Rate of Return Bands Incentive Regulation**	A rate of return band is set such that the lower limit is the allowed rate of return used in revenue requirements calculations. Earnings between the upper and lower limits are presumed due to efficiency enhancements and shared with customers in the form of temporary rate reductions. Earnings in excess of the upper limit are refunded. A variant is no upper limit, but a graduated sharing where customers' share increases as earnings increase.
C. **Price Bands**	Prices are allowed to vary between upper and lower limits of a specified price band. A variant includes "bands" with no lower limit.

D. Monopoly Services Regulation	Rate of return regulation is selectively applied only to those services (or firms) that are classified as non-competitive, non-basic or natural monopoly services.
II. MAJOR REGULATORY CHANGES	Alternatives that abandon or radically alter traditional rate of return regulation.
A. Revenue Regulation Only	Revenue requirements are not used to set individual prices, but functions only as a total revenue cap.
B. Price Regulation Only	Only the prices of individual services are regulated, and earnings or revenue regulation is abandoned.
1. Price Caps	The average price of a collection of services (price basket) is constrained to be less than some price cap. The price cap is automatically adjusted according to changes in inflation, gross productivity and exogenous cost changes.
2. Price Standards	Individual service prices are set according to one of a variety of economic price standards.
C. Sanctions/Deregulation	No price or revenue regulation is applied. Conformity with various public policies is assured by the application of (or threat of the application of) various criminal and civil sanctions, including re-regulation.
D. Public Ownership	Monopoly services, such as local telephone service, are owned and operated by local or state governments. Otherwise, there would be no regulation of prices, earnings or revenues.
E. Franchise Bidding	Regulated firms would be required to periodically bid for the right to provide monopoly service to a specified area. Money collected from the bid would be used to offset prices or advance various social goals. Otherwise, there would be no regulation of prices, earnings or revenues.

STATE REGULATORS, LEGISLATORS, AND U S WEST

by

Paul E. Teske

STATE UNIVERSITY OF NEW YORK AT STONY BROOK

Presented to the Sixteenth Annual
Telecommunications Policy Research Conference

Airlie, Virginia
October 31, 1988

ABSTRACT

STATE REGULATORS, LEGISLATORS, AND U S WEST

U S West has pursued the most aggressive strategy to achieve deregulation of all the Baby Bell siblings since the AT&T divestiture, at the federal level and particularly at the state level. The paper shows that this strategy was not motivated simply by "cowboy ideology"; U S West executives correctly perceived that the environment in which they operated differed from that in the rest of the country and that a different deregulatory strategy was appropriate. The firm exploited institutional differences in their states by bypassing state regulators and going directly to state legislators to get favorable policies. While they were sometimes too aggressive with this strategy, it has worked well, given the environment. This paper presents quantitative and qualitative evidence supporting the fact that institutional differences across states affect regulatory decisions and that U S West understood this and acted accordingly.

INTRODUCTION

From capture theory in the political science literature of the 1950s (Huntington, 1952; Bernstein, 1955) to economic theories of regulation in the 1970s (Stigler, 1971; Posner, 1974; Peltzman, 1976), scholars viewed the regulated firms as dominating regulatory decision-making. The implementation of deregulation in the 1980s either disconfirmed this simple theory of industry dominance (Wilson, 1980) or itself resulted partly from politicians understanding the accuracy of the theory in past regulatory practice (Derthick and Quirk, 1985).

What is the role of partially regulated firms in a deregulatory environment? Given the multitude of influential actors now participating in regulatory debates, including government-funded consumer groups, organized large users, and actual and potential competitors, how much influence do the regulated firms have with their regulators?

State telecommunications regulation after divestiture offers an excellent laboratory to compare the differential impact of interest groups on decision-making. The federal court-mandated divestiture was an exogenous shock to state regulators' stable environment. Suddenly they faced complex and important decisions about rate structures and entry into intrastate markets.

Since the 1984 creation of seven regional holding companies (RHCs), or "Baby Bells", in the Modification of Final Judgment (MFJ), the actions of one RHC have stood apart. U S West's Chairman said, "Divestiture created seven new companies. One of them isn't like all the others." (Vietor, 1986). U S West pressed for deregulation to a far greater degree than its siblings, and has taken a qualitatively different approach to achieving its goal. This paper explores that strategy, its effectiveness, using both quantitative and qualitative evidence, and places it in the context of business lobbying in regulatory policy.

THE CREATION OF U S WEST

The MFJ attempted to create viable local Bell operating carriers under a multi-state holding company structure. In the forty-nine continental states, including the District of Columbia, the seven RHCs averaged seven states and a population of about 33 million served. U S West was given fourteen western states (not California or Nevada) with sparse populations and few large cities. U S West became the parent company of Mountain Bell, Northwestern Bell, and Pacific Northwestern Bell.

One often overlooked reason that U S West executives planned a different regulatory strategy was the different environment in which they operated. Table 1 compares important characteristics of U S West's regulatory environment with national averages faced by the other six RHCs.

TABLE 1

Comparison of Average Values for variables in quantitative model -

14 U S West States versus 35 Non-U S West States

(State 1984 averages, not weighted by state population)

	U S West	Other States
Contextual Variables		
NECA Subsidy	+ $19 mill./st.	$-7.5 mill./st.
CATV Penetration	49%	46%
Access Loop Costs	$ 27.8/month	$ 25.2
Interest Groups		
* Fortune Serv HQs	3 / state	11.6 / state
Govt.-Funded Adv.	28% high	60% high
Non-Govt. Adv.	43% high	57% high
** U S West	All	None
Institutional Variables		
** Legisl. Party Control	57% Rep, 21% Mix	14% Rep., 20% M
Elected Commission	43%	23%
Reg.Climate (from 1-9)	4.5	4.75
* Regulatory Budget	$5.1 mill./st.	$9.6 mill./st.

* difference significant at 95%.
** difference significant at 99%.

The contextual climate differed in U S West states compared to that in other states, although none of the three variables are different at high levels of statistical significance. The NECA interstate subsidy averaged + $19 million for U S West states and was positive for all but Utah, while 25 of the other 35 states were sources of subsidies, for an average of - $7.5 million for all 35 states. The penetration percentage of cable television was slightly higher in U S West states compared to the other 35 states. Average local access loop costs were about two and one-half dollars higher in U S West states because of the lower population density.

U S West states contained fewer large service business headquarters, with an average of three Fortune Service 450 headquarters, as compared to 11.6 in the other states. Thus, this measure of the threat of service bypass by large users was significantly lower for U S West. Consumer advocacy in regulatory proceedings was less in U S West states (using the bivariate measures from Gormley, 1983). Government-funded advocacy activity was rated "high" in only 28% of their states (4 of 14), compared to 60% of the other 35 states (21 of 35); grassroots consumer advocacy activity was rated high in 43% (6 of 14) of U S West states, compared to 57% of the other states (20 of 35). The difference in consumer advocacy activity funded by government was significant at the 90% level.

While the external climate of interest groups and contextual variables differed somewhat for U S West, the institutional variables relating to the regulatory commissions and the legislatures in their fourteen states were quite different. Both houses of their state legislatures were much more often in the control of Republicans, some 57%, while only 21% were controlled by Democrats. In contrast, in the other 35 states, only [illegible] of state legislatures were completely controlled by Republicans, while 66% were controlled by Democrats. This difference is statistically significant at the highest levels.

A large percentage of U S West states featured directly elected utility commissioners, 43%, while in the other states only 23% were elected and 77% were appointed to the office. The regulatory climate, as measured by a Wall Street firm (Regulatory Research Associates), was rated somewhat lower in U S West states, 4.5, than in the other 35, where it averaged 4.75. The regulatory commission staffs in U S West states were smaller and had budgets of about one-half the average in the other 35 states. This difference, which was likely to affect the analysis of regulatory policy, is significant at the 95% level.

To the extent that these contextual, interest group, and institutional variables were expected to influence state regulatory decision-making (and my quantitative evidence presented here suggests that several of them were important), U S West faced a different decision-making environment than its other Baby Bell siblings. Contrary to popular belief, it was not high average access loop costs that differed significantly for U S West; while the company operates in less dense states, most of the people and businesses it serves are located in metropolitan areas. The institutional differences are far more important; U S West faced a better legislative climate for deregulation and a less favorable public service commission climate. I will show how U S West's strategy to deal with this environment was rational and successful to some degree.

Other evidence supports the argument about the more negative public service commission regulatory regulatory climate faced by U S West. Moody's Investors, Inc. prepared a report in 1983 on the likely prospects for the different local telephone companies after divestiture. For Mountain Bell, with one-half of U S West assets, serving Arizona, Colorado, Idaho, Montana, New Mexico, Utah and Wyoming, they cited a commodities-based economy, high population growth and large capital requirements. "Overall regulation of Mountain Bell is less than favorable. . . . Because of the great potential for intraLATA bypass, the company will need the regulators to reduce intraLATA rates" (p.3360). For Northwestern Bell, with 28% of U S West assets, serving Iowa, Minnesota, Nebraska, North Dakota and South Dakota, they noted, "With the exception of Nebraska, the regulation confronting Northwestern Bell has been relatively supportive. While we do not expect the consumerism of Nebraska to spread, nor do we anticipate a full appreciation by regulators of the risks now facing the telecommunications industry." (p.3362). For Pacific Northwestern Bell, serving Oregon and Washington, they said, "We consider the regulation for PNWB below average" (p.3364). Thus, the views of stock analysts confirmed that the regulatory climate for U S West was below average.

U S WEST'S STRATEGY

Clearly U S West faced a different regulatory decision-making environment than the other RHCs. Much anecdotal evidence suggests that U S West took a different approach to dealing with its regulatory environment, as well. The September 24, 1987 Wall Street Journal featured a front-page article on U S West and its high-pressure regulatory approach:

> "Almost since its creation in January 1984, U S West has been arguing strongly for total deregulation of its prices and profits in the 14 Western states where it operates . . None (of the other six RHCs) is seeking deregulation on the scale that U S West wants or has so irked regulators, legislators and consumer groups . . . 'If you don't make dust, you eat dust' is a favorite company slogan, and it typifies the rootin' tootin' cowboy image that the company likes to project. . . U S West is especially concerned about the growth of private networks. It says there are more than 1700 networks, mostly microwave systems, in its territory." (Roberts, 1987).

The December 2, 1985 Business Week issue discussed U S West and its Chairman, Jack MacAllister:

> "Some Bell executives, led by MacAllister, in essence are calling for the divestiture accord to be torn up . . . He has turned it into the most iconoclastic of the Bell siblings. . . MacAllister constantly rails against the evils of monopoly and regulation. . . MacAllister advocates wholesale deregulation of nearly every phone service in U S West's 14 state region" (p.96).

Harvard Business School researchers used U S West's reorganization strategy as a case study (Vietor, 1986). The strategy had three major elements: 1) separate competitive from regulated entities; 2) diversify into non-telecommunications activities; and 3) aggressively seek deregulation and competition in all markets.

As part of the separation strategy, U S West continually filed for waivers of federal regulation from Judge Harold Greene, who presided over the MFJ. Furthermore, U S West challenged in court whether the federal government's restrictions from the MFJ were applicable to it since U S West had not been a party to the original Consent Decree (it did not yet exist as an entity). While this specific challenge failed, U S West was aggressive consistently in trying to reduce, relax, and even eliminate federal regulation.

U S West sought to decentralize its operations: "While most other regionals control operations tightly from headquarters, U S West is set up as a decentralized holding company. Its three local phone companies and half-dozen new subsidiaries are run as separate entities" (Business Week, 1985).

State regulators found this organizational strategy difficult to manage. Sharon Nelson, Washington state regulator, characterized problems with U S West's constant reorganizations and the double holding company structure they adopted that serves no obvious purpose: "They create a shell under which to hide the pea." (Teske, 1987).

Ron Lehr, Colorado regulator, cited his concerns with the diversification of U S West:

> "One, cross-subsidies, including management attention, employee transfers, affiliate transactions, and capital costs; two, diversification motivations; three, an adequate regulatory response in small states . . . Yellow Pages were taken away unilaterally by U S West . . . It was like the black ship sailing away flying a black pirate flag, while I'm rowing a dinghy of regulation." (Teske, 1987).

To develop the second element of the strategy, U S West established new subsidiaries and acquired other firms. They began a consulting firm called Interline Communications, a real estate company called BetaWest, and a firm selling cellular networks in the Mexican gulf called New Vector Communications. They purchased Acquired Commercial Funding, a $10 million financial services company, Trans Western Publishing, and Control Data's Kansas City Commercial Credit Company. Business Week eported that, "By 1990 he (MacAllister) hopes to get 50% of its (U S West's) revenues from unregulated businesses, up from 10% this year" (p.97).

The third element of the strategy is most relevant to this study. U S West's general counsel, Larry DeMuth, stated bluntly, "Regulation is irrelevant in an environment of competition." (Vietor, pg. 147), and, "The user is going to simply ignore regulation and develop around it . . . it's irrelevant today" (Nadel, 1986). Other Bell regionals understood U S West's approach, "Says a top official at another regional - "Their philosophy is that competition is good, therefore do it."" (Business Week, p. 97).

U S West pushed for deregulation and competition at all levels of government. But, as the Wall Street Journal noted, ". . . although federal regulation gets more attention, state regulation of the Baby Bells currently affects them more strongly . . .it is at the state level that U S West's deregulation campaign stands out for breadth and controversy." (Roberts, 1987). At the state level, U S West's regulatory strategy was clear; avoid state public utility commissions and seek deregulation by state legislators. If approved legislation is the measure of success, the strategy worked well; of the 18 state legislative deregulatory bills passed from 1984 through the end of 1987, ten were passed in the fourteen U S West states.

U S West used its political clout and economic development promises to achieve this success. According to the Wall Street Journal:

> "In Washington, a state official says the company's Pacific Northwest Bell unit told her it probably wouldn't locate a big new research center in the state because of her pro-regulation views. . . 'Linking economic development with deregulation of monopoly telephone service is inappropriate', she (Sharon Nelson) argues. . . Officials in Arizona and Iowa complained that U S West tried to link the site selection to its deregulation campaign. . . In Colorado, U S West's Mountain Bell unit has publicly vowed to throw its considerable political weight against any lawmaker who votes to renew the state's consumer advocate post, now held by a man often critical of the company's deregulation and pricing policies. . . In Colorado, when the deregulation bill was being considered, the company's Mountain Bell Foundation pledged some $100,000 for a project in a legislator's district; the company presented the gift to her within hours of a meeting at which she discussed the bill with Mountain Bell employees." (Roberts, 1987).

QUANTITATIVE MODEL

In my quantitative model I attempt to explain and predict state regulatory choices in a cross-sectional analysis during 1984-1986. The dependent variables explained are state rate structure decisions and state intraLATA competition choices. (LATAs are local access and transport areas, generally corresponding to metropolitan areas and/or telephone area codes, and were established as the likely area of "natural monopoly" or limited competition by the MFJ.)

While there are many potential specifications of rate structure policy, I have chosen to measure it as movement towards efficient economic pricing. If a state either lowered toll rates within its boundaries and/or established an unbundled local access charge, it is assigned a 1 for this dependent variable. Otherwise it is assigned a zero. This criteria exactly mirrors FCC policy since the MFJ that has reduced interstate toll rates (by about 35%), while establishing a customer access line charge, or CALC ($2.60 per month in 1988), on all residential bills. Nearly one-third (16) of the 49 states changed rate structures between 1984 and 1986.

For competition, a state is assigned a 1 if it has allowed intraLATA competition and a zero otherwise. IntraLATA toll calls represent between one-quarter and one-third of the national toll market, and the decision to allow or disallow competition in this market is of great importance, particularly to Baby Bell competitors. Fifteen states opened intraLATA markets to competition between 1984 and 1987.

My regression estimates use the logit model, appropriate for a binary dependent variable. Included as independent variables are the measures presented in Table 1. These include contextual variables that capture important elements of the economic environment, interest group variables that capture the strength of organizations that can pressure regulators for policy change, and institutional variables that reflect differences across states in regulatory structures and attitudes. The general hypothesis is that all of these factors may influence regulatory policy.

Economists and political scientists have used versions of interest group theory to explain regulatory policy (for example Stigler, 1971; Peltzman, 1976; Owen and Braeutigam, 1978; Ackerman and Hassler, 1981; Noll and Owen, 1983). Other scholars have stressed the importance of institutional factors, including legislators, regulators, and regulatory bureaucracies (Wilson, 1980; Kalt, 1981; Derthick and Quirk, 1985; Sabatier, 1987).

Specific hypotheses are presented in Table 2.

TABLE 2

HYPOTHESES ABOUT INDEPENDENT VARIABLE IMPACTS ON

	RATE STRUCTURE	COMPETITION
Contextual Variables		
NECA Subsidy	-	-
CATV Penetration	-	+
Access Loop Costs	+	-
Interest Group Variables		
Fort Service 450 HQs	+	+
Govt-Funded Advocates	-	+
Non-Govt Advocates	-	+
US West	+	+
Institutional Variables		
Rep. Leg. Party Control	+	-
Elected Commission	-	+
Regulatory Climate	+	+
Regulatory Budget	+	+

The data come from the National Association of Regulatory Utility Commissions, the 1986 State and Metropolitan Area Data Book, Regulatory Research Associates, the FCC, the National Exchange Carriers Association (NECA), and other industry sources. The data are all from 1984 or 1985. The models themselves are described more fully in Teske, forthcoming.

The results of the regressions show important relationships between several of the independent variables and rate structure and competition decisions. Table 3 shows the results for rate structures choices.

TABLE 3

EFFECT OF INDEPENDENT VARIABLES ON RATE STRUCTURE DECISIONS

	Beta	Std. Error	Significance
Contextual Var.			
NECA	-0.02	0.01	89%
CATV	-0.04	0.05	60%
Loop Costs	0.01	0.02	55%
Interest Groups			
Fort 450 Serv HQs	-0.09	0.10	66%
Govt-Fund. Adv.	0.82	0.94	62%
Non-Govt. Adv.	0.20	0.88	18%
U S West	0.94	1.05	63%
Institutional Var.			
Leg. Party Control	1.04	0.58	92% *
Elected Comm.	1.36	1.09	79%
Reg. Budget	0.31	0.14	97% **
Reg. Climate	0.39	0.20	95% **

N = 49

Logit Regression Chi-Square = 22.7 (98%)

The strongest variables are the budget, the regulatory climate, and the legislative party control. The dummy variable for states served by U S West is not significant, but is positively related to rate structure changes. Much of its explanatory power is taken away by the Republican legislative party control, with which it is correlated at a level of .46.

I tested for the possible endogeneity of the regulatory budget and regulatory climate variables, using a two-stage model, and I found that the budget continued to have strong positive effects of a similar magnitude. I was unable to find proper instrumental variables to "exclude" the regulatory climate in a first-stage regression, so I must conclude that it has an independent, positive impact on pricing choices as well.

Table 4 shows the sensitivity analysis for the variables with the largest impacts. With all other variables evaluated at their means, the impact of a change in regulatory climate from one standard deviation below its mean (SDBM) to one above (SDAM), or from 1.9 to 7.5, is a 38% increase in the probability that a commission will change the rate structure.

TABLE 4

MARGINAL EFFECT OF VARIABLE ON RATE STRUCTURE DECISIONS
WITH OTHERS SET AT MEAN

	SDBM	SDAM	Change
Contextual Var.			
NECA	.53	.08	-.45
CATV	.32	.17	-.15
Loop Costs	.17	.32	.15
Interest Groups			
Fort 450 Serv HQs	.41	.09	-.32
Govt-Fund. Adv.	.22	.25	.03
Non-Govt. Adv.	.17	.31	.14
U S West	.19	.32	.13
Institutional Var.			
Leg. Party Control	.13	.43	.30
Elected Comm.	.17	.37	.20
Reg. Budget	.03	.85	.82
Reg. Climate	.10	.48	.38

The variables with the largest impacts are the regulatory budget, the NECA interstate subsidies, the regulatory climate, the Fortune Service 450 headquarters, and the legislative party control. The impact of U S West states versus non-U S West states is a 19% increase in the probability of a decision to change prices (the table shows the standardized difference (.13); 19% is the impact for a full binary change).

Using these parameters for each state, the model predicts very well. Forty-two of the 49 cases are predicted correctly (where a model result of .5 or more and an actual policy change is correct, and .49 or less and no change is correct), for 86% accuracy. The model thus reduces the error in prediction from simply using the modal value by over 60%. All U S West states are predicted correctly, with the exception of Wyoming, which is not predicted to change rate structures but which actually did so. In contrast, an alternate specification of the model using only contextual and interest group variables is not as strong; only 69% of the cases are predictly correctly, barely better than the modal value of 67% "no change" in pricing policy.

The model with only contextual and interest group variables does confirm the importance of U S West as an independent actor. Table 5 shows the results; U S West states are significantly more likely to change prices.

TABLE 5

EFFECT OF CONTEXTUAL AND INTEREST GROUPS VARIABLES ON RATE STRUCTURE DECISIONS

	Beta	Std. Error	Significance
Contextual Var.			
NECA	-0.01	0.01	79%
CATV	-0.03	0.04	54%
Loop Costs	0.01	0.01	50%
Interest Groups			
Fort 450 Serv HQs	0.08	0.04	95% **
Govt-Fund. Adv.	0.13	0.76	14%
Non-Govt Adv.	0.11	0.73	12%
U S West	1.54	0.87	92% *

N=49
Logit Regression Chi-Square = 8.6 (71%)

This "external" model shows that U S West is highly significant when the Republican legislative variable is excluded. Large users, as represented by Fortune Service 450 headquarters are also highly significant in Table 5, when the regulatory budget, with which it is highly correlated (.82) is left out.

Table 6 shows the results of the fully-specified model for the competition choice. Again several variables show an important relationship to the competition choice.

TABLE 6

EFFECT OF INDEPENDENT VARIABLES ON COMPETITION DECISIONS

	Beta	Std. Error	Significance
Contextual Var.			
NECA	-0.01	0.01	54%
CATV	0.06	0.07	62%
Loop Costs	-0.01	0.03	34%
Interest Groups			
Fort 450 Serv HQs	0.15	0.12	80%
Govt-Fund. Advocates	5.63	2.74	96% **
Non-Govt. Adv.	-0.49	1.44	26%
U S West	8.40	3.92	97% **
Institutional Var.			
Leg. Party Control	-3.53	1.49	98% **
Elected Comm.	-4.56	3.05	87%
Reg. Budget	-0.08	0.10	55%
Reg. Climate	0.55	0.34	89%

N = 49
Logit Regression Chi-Squared = 38 (99%)

In this model, despite the importance of legislative party control, the independent impact of U S West on competition choices is strong, significant, and positive. Other variables that are important are government-funded consumer advocates, large users, elected commissioners, and regulatory climate. This model predicts state choices very well; 44 of the 49 cases are predicted correctly, for 90% accuracy, a 68% reduction in the error of prediction using the modal value.

Table 7 shows the marginal impact of variables, with other held at their means, and the importance of U S West is established.

TABLE 7

MARGINAL EFFECT ON COMPETITION DECISIONS OF INDEPENDENT VARIABLE WITH OTHERS SET AT MEAN

	SDBM	SDAM	Change
Contextual Var.			
NECA	.15	.05	-.10
CATV	.05	.15	.10
Loop Costs	.13	.06	-.07
Interest Groups			
Fort 450 Serv HQs	.02	.41	.39
Govt-Fund. Adv.	.01	.60	.59
Non-Govt. Adv.	.11	.07	-.04
U S West	.01	.82	.81
Institutional Var.			
Leg. Party Control	.56	.01	-.55
Elected Comm.	.26	.01	-.25
Reg. Budget	.16	.04	-.12
Reg. Climate	.02	.31	.29

The movement from a standard deviation below the mean to a standard deviation above the mean for U S West leads to a greater increase in the probability of regulators approving competition than for any other variable. The difference between U S West and other states, with other variables set at their means, is a 96% greater likelihood of competition in U S West states (81% is the standardized difference).

Thus, quantitative evidence confirms that institutional variables have important effects on state rate structure and competition decisions. Even with these variables controlled for, U S West states are far more likely to allow competition and more likely to move towards efficient rate structures. The case studies show how this happened in a few specific states.

CASE STUDIES

U S West's main strategy for state deregulation was to avoid regulators and try to influence legislators to pass bills forcing deregulation. The quantitative evidence supports this strategy by showing that the U S West dummy variable and legislative party control had strong and significant impacts on state policy choices. How was U S West able to pursue this strategy?

I believe U S West successfully pursued this strategy because state legislators have incentives to care more than regulators about economic development and jobs in their state. During much of the post-divestiture period, the commodities-based economies of the U S West region performed poorly. U S West used this concern when it attempted to tie the economic development it controlled to regulatory outcomes.

In Nebraska, U S West succeeded. In Arizona, they succeeded with the legislature but a state constitutional requirement sent them to the voters. In Idaho, they succeeded with the legislature, but were partially foiled by the governor. U S West did not limit its political pressure to these three states; by mid-1988 they had successfully achieved deregulatory legislation in 11 of their 14 states. Many of these laws had a three-tiered approach to deregulation. The Wall Street Journal noted: " . . other states in U S West's territory, including Minnesota and Iowa, totally deregulated some services, singled out some for possible relief, and retained rules on others, such as the local phone calls that U S West monopolizes". (Roberts, 1987).

Nebraska

Politicians in Nebraska adopted a different approach to telephone deregulation than all other states. In 1986, the legislature passed and the Governor signed a law to allow radical price deregulation of all services, including local services, which was the most extreme reduction in regulation in the nation.

Northwestern Bell, the major local telephone company serving Nebraska, is part of U S West. The legislation, written by U S West, allowed local telephone companies to increase local prices up to 10% annually, with 90 days notice, unless 2% of affected consumers signed a petition opposing the increase, in which case regulators would hold hearings to make a determination. Services other than the local exchange were nearly freed of any rate-of-return regulation.

The law retained some state regulatory control over telephone companies for a five-year period. Local service quality and entry/exit were strictly regulated. The law did not allow open entry into the local exchange market. IntraLATA toll competition was not authorized until 1989 and interLATA rate de-averaging was prohibited until 1991. After 1992, however, telephone companies in Nebraska will be free to set all rates as the market will bear.

The legislature enacted the law despite the strong objections of the public service commissioners. Commission President Harold Simpson said, "How do you force good service if you don't have a handle on rates?" (Davis, 1986). The new state Attorney General sued to have the law declared unconstitutional. In May 1987, a Lancaster County District Court ruled the law, LB 835, "constitutional and valid". The opposition appealed the ruling to the State Supreme Court, but the law was implemented in the meantime. Three Senate bills attempted to reverse the deregulatory legislation.

U S West used legislation to get around regulators who were reluctant to implement large local access increases in the rural western states. While Nebraska regulators did not give Northwestern Bell all that it requested, they were more generous than regulators in several neighboring states. In a 1984 rate case, Nebraska commissioners gave Northwestern Bell a $22.6 million increase, of which three-quarters went to local price hikes. Local flat rates rose from $11.64 to $14.14 in Omaha. In a study of rates in 17 western states, roughly corresponding to the U S West service area, residential local rates in Nebraska were only 12th highest in 1983 but were fourth highest in 1985, because Nebraska's rates rose more than in many neighboring states (Energy and Resources Consultants, Inc, 1986). Business rates increased substantially as well, with only New Mexico, among these western states, implementing larger business price increases from 1983 to 1985.

By 1986, however, Nebraska regulators were less generous. They chose not to approve any of a proposed rate hike, although by law their inaction on the case caused an automatic partial increase. Prior to the legislation Nebraska regulators were not innovative in other aspects of telecommunications regulation. The commissioners resisted intraLATA competition originally, but since the passage of LB 835, they approved competitive petitions by MCI, U.S. Sprint, and AT&T to provide intraLATA service.

Why did legislators pass such a radical bill, with regulators opposed, when the commission had seemed more favorable to gradual deregulation than those in several neighboring states? Nebraska has the only unicameral, officially non-partisan-elected legislature in the nation. As a result, U S West needed approval from only one legislative body, which may have aided passage of the bill. As in other states, U S West used its full political clout, including threats to shift jobs to other states, to get the bill passed. It appears that such a radical bill passed in Nebraska largely because former Democratic Governor Robert Kerrey strongly supported it and, as lead telecommunications spokesperson for the National Governor's Association, sold the bill as markedly enhancing Nebraska's attractiveness to high-technology, telecommunications-dependent firms, as part of an economic development strategy. Nebraska had a head-start in this area; the presence of the Strategic Air Command and several hotel chain reservation systems in the state showed success with telecommunications-dependent operations.

Furthermore, Nebraska had been hit particularly hard by agricultural failures and almost any form of potential economic development seemed desirable. This economic climate (although it was partially shared by neighboring states) and the support of the governor probably account for the bill's passage. As a success of the bill, Kerrey cited a telemarketing firm that established operations in Nebraska in 1986 to handle over one million calls for Kodak when it stopped producing its instant camera. He claimed that 12 to 15 other firms moved to Nebraska as a result of the relaxed regulatory climate. Kerrey's successor, Republican Governor Kay Orr, continued this general economic development strategy by offering telecommunications and data processing firms substantial tax credits for creating jobs and expanding in Nebraska.

The following quotes illustrate former Governor Kerrey's views on this issue: "You had U S West making the case for this all over its region, and here in Nebraska they had a receptive audience . . . If they don't have to worry about some arbitrary commission of Nebraskans setting their prices, they'll come in here and show us what all their technology can do for our people . . . If you live in a rural, isolated state like Nebraska, you absolutely need to be connected to the rest of the country. And there is technology coming along that can connect us much more closely. But to get it, we have to move away from arguing, 'What should the price of the product be?' and into 'What should the product be?'" (Reid, 1986).

An irony first noted by Gene Kimmelman of the Consumers Federation of America, was that Nebraska's "first mover" advantage may have protected the state since U S West wanted to sell this deregulatory concept to other states and it needed to show that it caused minimal political and economic harm (Roberts, 1986).

The aggressive push of U S West and the economic development agenda of the governors account for the success of legislative deregulation. The elected public service commissioners in Nebraska did not agree or coordinate with legislative actions. Deregulation in Nebraska thus involved minimal public service commission input.

Arizona

A referendum in Arizona on November 4, 1986, Proposition 100, provided voters with their first chance to make a direct statement about state level deregulation of telephone services. Referendum instructions advised voters that a "yes" vote "would authorize the Corporation Commission to reduce regulation of some telecommunications services, and assure statewide availability and affordability of telephone services using definitions and procedures prescribed by legislation." Voting "no" would "require the Corporation Commission to continue to regulate all telephone and telegraph services" (Megdal, 1986). The referendum was necessary because the Corporation Commission's (ACC) power is delineated specifically in the state constitution, unlike most other states. Proposition 100 effectively would have passed much regulatory authority from the Corporation Commission to the legislature. It was defeated by a 54-46% margin.

Mountain Bell, the U S West subsidiary, and AT&T spent $3 million, through a political action committee, to support Proposition 100. U S West was extremely interested in the referendum because, while Arizona totaled only 12% of its revenue, the state represented 38% of its recent growth (Megdal, 1986). As it did elsewhere, U S West played political hardball in lobbying for passage of the referendum; during the commission's consideration of Mountain Bell's 1986 rate increase request, for example, U S West threatened to move 1800 jobs to New Mexico if approval was not forthcoming.

Opponents of Proposition 100 spent approximately $50,000, and thus were outspent by about a 60:1 ratio. The opponents were well-organized, however, due in part to the efforts of the United Church of Christ, active telecommunications consumer advocates.

The message that voters sent by defeating the referendum was not completely clear. Mountain Bell suggested that is showed "customers expressing confidence in the Corporation Commission's regulation of telecommunications services" (NARUC Bulletin, 1986). My interpretation is that voters opposed further deregulation since initial state and federal policies hurt and confused them. Voters expected the regulators to move more slowly than the legislature because they viewed regulators as captured less by U S West.

Prior to the referendum, the two legislative houses passed conflicting bills. The Arizona House passed a Mountain Bell-sponsored bill to implement immediate deregulation leaving only the local loop for Corporation Commission regulation, almost as radical a bill as passed in Nebraska. The House bill was opposed by the state Residential Utility Consumer Office along with the Communications Workers of America, the commission, and competitors, but was nevertheless approved by a 2:1 margin. The Senate sought slower deregulation, with legislative control and the commissioners as referees. The Senate bill provided for service-by-service deregulation, a far less radical approach. While commission interpretation of rules would have continued, they were no longer to be the main policy makers (Megdal, 1986).

The Arizona commissioners were elected directly by voters and serve six year terms. This provided for more electoral independence from the legislature and helps to explain the conflict between the two institutions. Prior to the referendum, the commissioners moved towards economically efficient pricing by increasing local rates and reducing toll rates generally, as well as reducing rates on the four most dense intrastate calling routes. They did not approve intraLATA competition.

Thus, Arizona illustrates voter opposition to deregulation, despite the large expenditure by U S West. The legislature favored the U S West position but needed voter approval. The voters would not authorize increased legislative control over regulators despite the policies of regulators that had favored deregulation and efficient pricing more than those in most other western states.

Idaho

In the immediate post-divestiture period, Idaho regulators were among the most outspoken opponents of FCC policies favoring a shift of costs from toll rates to end-users. The public utility commission tried, unsuccessfully, to order Mountain Bell, the U S West subsidiary that serves almost three-quarters of the access lines in Idaho, to refund the 1985 $1 federal customer access line charge to residences that did not use toll services. Idaho regulators continued "value of service" pricing, in which households that can reach a larger number of other subscribers with a local call pay a higher local flat-rate than those in smaller local exchange areas. "Value of service" pricing contradicts pricing at economic costs, which leads to lower access rates for subscribers in larger urban areas.

The commission explicitly set a goal of non-traffic sensitive cost coverage at 50% local, 25% intrastate toll, and 25% interstate toll, and they ordered modest local rate increases towards this goal. Even when these goals are achieved, they still represent a large cross-subsidy from toll to local access services. In 1984 and 1985 Mountain Bell rate cases regulators spread increases to services across-the-board, at about 3%, rather than changing rate structures. As a result, from 1983 to 1985, average residential flat-rates rose from $9.07 to only $11.31 per month, less than the national average increase.

Idaho regulators were opposed to increasing local access charges directly. They also opposed other changes that would tend to raise local rates. Due to their finding of past over-earnings, regulators instituted a rate freeze for Mountain Bell for 1987 and most of 1988. When the U.S. Supreme Court's 1986 Louisiana ruling left them with discretion of this issue, Idaho regulators rescinded their previous approval of accelerated depreciation rates.

In 1987 U S West pushed the Idaho legislature to pass a deregulatory bill similar to bills it had lobbied for successfully in 9 of the other 13 states in which it operated. The legislature was much more favorably disposed to U S West than were the regulators. Governor Andrus surprised most analysts by vetoing Idaho bill HB 149 in 1987, and his veto was sustained. Although the House State Affairs Committee changed the specific elements of the bill to which the Governor objected, legislators were not able to reconsider the bill until they returned in 1988.

Andrus objected to a few specific elements of the bill, including a clause that removed commission authority over rate increases of less than $1 per month in a year and a clause that removed regulatory authority over customer billing complaints.

The bill also forced local exchange rates that were below the rates in the largest exchanges to increase each year by one-third of the difference, to eliminate "value of service" pricing. Andrus said the bill "places the universal availability of the telephone in jeopardy for too many of our citizens." (Roberts, 1987).

Thus, while Idaho regulators made some changes in telephone regulation, the state was near the extreme end of the fifty states in maintaining cross-subsidies and limiting competition. U S West attempted to use legislative action to compel deregulation, as succeeded in Nebraska but back-fired in Arizona. The governor's veto foiled this strategy in the short-run. A new version of the bill passed in 1988, but like a bill passed in Colorado recently, it was far less favorable to U S West, and some company executives feel that it was little better than no bill at all.

CONCLUSION

Some observers have characterized U S West's approach to deregulation as wild, aggressive, and extreme. Renz Jennings, a member of Arizona's Public Utilities Commission says, "U S West is pursuing a school-yard bully approach to deregulation" (Wall Street Journal, 1987). Others have argued that an aggressive approach was necessary due to the low-density, rural character of much of U S West's territory. These descriptions may be partially accurate, but they miss the rationality of U S West's political strategy, given relatively hostile regulators and relatively favorable legislators, as well as the other important differences in the regulated environment tested in my analysis. Although U S West may have pushed too hard, which lead to the loss of political support in some states, they succeeded in getting deregulatory bills passed.

Why have legislators been so much more favorable to U S West than regulators? First, legislators care more about economic development and other non-regulatory advantages to relaxing regulation. Second, voters have more trouble tracing negative impacts from legislative deregulatory bills, which focus on goals goals and policy directions, than they do for specific rate increases approved by regulators. Third, U S West is one of the largest employers in the region and was able to exercise raw political muscle more effectively than were opponents (a not insignificant proportion of part-time legislators in these states had some connection to telephone companies).

The efforts of U S West suggest that a single firm in a deregulatory environment can have a strong influence on the direction, degree, timing, and scope of deregulation. Despite the constellation of other interests in these states, U S West carried the most clout, as the traditional theories of regulation argued. Most of this clout came through economic development leverage and the political power held by the region's largest employer, with about 80,000 employees. Despite these levers, U S West was not able to get its way in every state. Different institutional and political factors in each state had independent effects on the extent of deregulation.

BIBLIOGRAPHY

Ackerman, Bruce and William Hassler, 1981. Clean Coal/Dirty Air. Yale University Press.

Bernstein, Marver. 1955. Regulating Business By Independent Commission. Princeton University Press.

Business Week, December 2, 1985. "Baby Bells Go Their Own Way".

Davis, Bob. 1986. "Many States Deregulate Telephone Rates, Hurting Residential Users in Short Run ", Wall Street Journal, September 19, 1986.

Derthick, Martha and Paul Quirk, 1985. The Politics of Deregulation. Brookings Institution.

Gormley, William. 1983. The Politics of Public Utility Regulation. Pittsburgh University Press.

Huntington, Samuel. 1952. "The Marasmus of the ICC", Yale Law Journal, April, p. 467.

Kalt, Joseph. 1981. The Economics and Politics of Oil Price Regulation. MIT Press.

Megdal, Sharon Bernstein. 1986. "The Political Economy of Telecommunications Deregulation in Arizona."

Moody's Investors, Inc. 1983. "Report on the AT&T Divestiture."

Nadel, Mark. 1986. "The Changing Mission of Telecommunications Regulators at the State Level", Aspen Institute Program on Communications and Society.

National Association of Regulatory Utility Commissioners (NARUC) Bulletin, 1986.

Noll, Roger and Bruce Owen, editors. 1983. The Political Economy of Deregulation. American Enterprise Institute.

Owen, Bruce and Ronald Braeutigam, 1978. The Regulation Game, Ballinger.

Peltzman, Sam. 1976. "Toward A More General Theory of Regulation", Journal of Law and Economics, p. 211.

Posner, Richard, 1974. "Theories of Economic Regulation," Bell Journal of Economics and Management Science, Autumn.

Reid, T.R. 1986. "Phone Deregulation, Phase 2", Washington Post, May 27, 1986.

Roberts, Johnnie. 1986. "Nebraska Law Deregulating Phone Rates May Spur Some Changes in Other States", Wall Street Journal, May 5, 1986.

Roberts, Johnnie. 1987. "U S West, Demanding Further Deregulation, Sparks Much Criticism", Wall Street Journal, September, 24, 1987.

Sabatier, Paul and Neil Pelkey. 1987. "Incorporating Multiple Actors and Guidance Instruments in to Models of Regulatory Policymaking: An Advocacy Coalition Framework". Administration and Society, August.

Stigler, George. 1971. "The Theory of Economic Regulation", Bell Journal of Economics and Management Science, p.3.

Teske, Paul. 1987. "State Telecommunications Regulation: Assessing Issues and Options in the Midst of Changing Circumstances, Aspen Institute Program on Communications and Society.

Vietor, Richard and Davis Dyer, ed. 1986. Telecommunications in Transition, Harvard Business School.

Wilson, James Q., editor. 1980. The Politics of Regulation, Basic Books.

FREQUENCY SPECTRUM DEREGULATION, PROPERTY RIGHTS AND MARKETS:
WHERE ARE WE NOW?

by

Douglas W. Webbink [1]

Federal Communications Commission
Mass Media Bureau
Policy and Rules Division
Washington, DC 20554

To Be Presented At:

The Sixteenth Annual Telecommunications
Policy Research Conference

Airlie House
Airlie, Virginia

November 1, 1988

---

[1] The views expressed are those of the author and do not necessarily reflect the views of the Commission. This is a slightly revised version of a paper presented at Columbia University in October, 1988. I wish to thank Steven Bailey, Judith Herman, James McNally, David Siddall and Rodney Small for helpful comments and suggestions. Opinions and errors remain my own.

TABLE OF CONTENTS

Page

I. Introduction 3

II. Recent FCC Spectrum Deregulatory Activities 9

  A. Spectrum Use Operational Flexibility 9

    1. Provision of Secondary Services 9

    2. Multiple Services In The Same Frequency Block 13

    3. Time Sharing 15

    4. Repeal of Permissible Use Restrictions 16

  B. Technical Flexibility 16

    1. Technical Flexibility And The Question of Mandating Technical Standards 17

      a. Mass Media Technical Flexibility 18

      b. Private Radio and Common Carrier Mobile Radio Technical Flexibility 20

      c. Other Flexible Technical Standards 22

  C. Frequency Coordination And Privately Negotiated Interference Agreements 22

    1. Frequency Coordination 23

    2. Privately Negotiated Interference Agreements 24

  D. Techniques For Expediting Spectrum Authorization: First Come, First Served, Lotteries, Auctions 27

    1. First Come, First Served 27

    2. Lotteries 28

Page

3. Auctions 31

E. Revenue Sources And Spectrum Pricing: License Fees, Spectrum Fees And Auctions 33

1. Cost Of Service License Fees 33

2. Spectrum Fees 34

3. Auctions 35

F. Other Restrictions on Spectrum Usage 36

1. License Term Length And Transferability 36

2. Petitions To Deny And Economic Impact: The Carroll Doctrine And UHF Impact Statements 38

III. Establishing Full Property Rights And The Need To Revise The Communications Act of 1934 38

IV. Why Incumbent Licensees Often Oppose Spectrum Deregulation And Spectrum Markets 41

V. Conclusions 44

I. Introduction

In the last 8 or 10 years, the FCC has attempted in a variety of ways to increase the play of market forces in the process of allocating and assigning the right to use the radio frequency spectrum. The FCC has also moved to give licensees more technical and operational freedom to decide how to use their licenses, once a license is granted. At the same time, the FCC has strived to decrease its own level of command and control over certain kinds of spectrum utilization decisions. Much of the impetus for these changes has come from the writing of outside academic scholars over the years and from the FCC's own staff. The interest of those writers has, in turn been caused in part by the increasing spectrum congestion in major urban areas due to the development of new technologies and new services that use spectrum. In many cases representatives of major radio licensee groups have been notably unenthusiastic, at least initially, about embracing the spectrum deregulation, spectrum flexibility, spectrum market and spectrum property rights ideas.

In 1980, I wrote an FCC Office of Plans and Policy Working Paper on spectrum deregulation ideas. [2] This present paper will review some of the major recent deregulatory decisions the FCC has made with respect to spectrum use and some of the proposals that are still under consideration. The paper will particularly emphasize actions the Commission has taken since 1980. While the paper will describe many of the major proposals and decisions that might be described as encouraging spectrum deregulation or the development of more complete spectrum markets; nevertheless, there are many additional decisions and proposals that have not been described. [3] This paper will also comment on this author's interpretation of the significance of some of these decisions.

---

[2] Douglas W. Webbink, "Frequency Spectrum Deregulation Alternatives," FCC Office of Plans and Policy Working Paper No. 2 (October 1980).

[3] Many of these spectrum deregulatory activities have taken place in combination with other Commission deregulatory actions in the Common Carrier, Mass Media and Private Radio Services. A quick and useful review of major FCC rulemakings is provided each year in the FCC Annual Report. For additional surveys of recent FCC deregulatory activities, see: Douglas W. Webbink, "Radio Licenses and Frequency Spectrum Use Property Rights," Communications and The Law, vol 9 (June 1987), pp. 3-29; Jill Abeshouse Stern, Erwin G. Krasnow and R. Michael Senkowski, "The New Video Marketplace and the Search for a Coherent Regulatory Philosophy," Catholic University Law Review, vol 32 (Spring 1983), pp. 529-602; Douglas W. Webbink, "The Recent Deregulatory Movement At The FCC," in Telecommunications In The U.S.: Trends

For at least the last twenty years, various writers have been pointing out the problems with the existing spectrum regulatory scheme. Because the Commission makes the basic allocation and reallocation decisions that determine overall allowed uses of the spectrum through an administrative process that is subject to all the restrictions and requirements of the Administrative Procedure Act, it may take years to implement those changes even when most parties agree to the proposed changes. When major interested parties disagree concerning the proper course of action, it may take far longer. Docket 18262, which reallocated UHF TV channels 70-83 (806-890 MHz) to cellular radio and private land mobile use was opened in 1968 and actually made the major reallocation decisions in 1974. [4] The Commission began accepting applications for cellular licenses in the largest 60 markets only in 1982, and it was 1983 before the first regular cellular system was authorized. By 1988, cellular systems had finally been constructed in most of the largest urban centers, but the FCC still has not authorized cellular service in many rural areas outside of the 305 largest metropolitan statistical areas. This may be an extreme example, but it is certainly not the only case of regulatory changes which take many years to implement.

In addition, generally allocations are made on a nationwide basis and hence may not reflect local differences in spectrum demand and availability. Most of the spectrum is divided into blocks that may be used only for specific and limited purposes. For example, part of the spectrum is allocated for land mobile communications and then subdivided so that it may be used only for public safety (police, fire, emergency medical) purposes. Indeed, the spectrum is divided into over 430 separate blocks by the FCC allocation table. [5] Moreover, the Commission often finds it very difficult to determine the most valuable use of the spectrum, especially when there are possible new but untested alternative uses for the spectrum. Users are restricted in their ability to buy, sell, resell, subdivide and combine blocks of spectrum. Because users do not pay explicitly for the right to use the spectrum, they do not necessarily take into account its value in terms of alternative uses (i.e. its opportunity cost)

And Policies, edited by Leonard Lewin, (Dedham, Mass.: Artech House, Inc., 1981), pp. 61-83.

[4] See: Notice of Proposed Rule Making and Notice of Inquiry in Docket No. 18262, 33 FR 10807 (July 17, 1968). See also: Second Report and Order in Docket No. 18262, 46 FCC 2d 752 (1974).

[5] 47 CFR 2.106 (1986).

and thus are unlikely to use it as efficiently as they would if they directly paid the opportunity cost for its use. In addition, because users (i.e. licensees) are not given full property rights in spectrum use, they may be unable to increase their profits by adopting various spectrum saving or spectrum using techniques. As a result, they do not necessarily have strong incentives to invest in new innovative techniques. For example, if users develop technologies that split their channels in half, they may be doubtful about whether the FCC will take half of the newly available channels away from them. In that case, those users may choose not to use or to reveal the existence of the channel splitting technique. [6] Finally, in the past, when the Commission received competing applications for a particular licensee, it often used a comparative hearing to decide who should receive the license. Comparative hearings often take years and do not necessarily lead to the award of the license to the person most likely to provide the service desired by consumers or businessmen.

For all these reasons, many writers have been critical of the FCC allocation and assignment process and have suggested that it leads to inefficient use of the spectrum. As a consequence, these critics have proposed a variety of alternatives. As early as 1959, Ronald Coase suggested establishing property rights in the spectrum and then auctioning off those rights to private parties. [7] Harvey Levin has been writing about improving the economic efficiency of spectrum allocation and use decisions since 1966. [8] Others have been writing about alternatives to the existing spectrum regulatory system since the 1960's. [9]

---

[6] See, for example, Douglas W. Webbink, "Frequency Spectrum Deregulation Alternatives," FCC Office of Plans and Policy Working Paper No. 2 (October 1980).

[7] Ronald H. Coase, "The Federal Communications Commission," Journal of Law and Economics, vol II (October 1959), pp. 21-40.

[8] Harvey J. Levin, "New Technology and the Old Regulation in Radio Spectrum Management," American Economic Review Papers and Proceedings, vol. 56 (May 1966), pp. 339-346; Harvey J. Levin, "The Radio Spectrum Resource," Journal of Law and Economics, vol. XI (1968), pp 433-501; Harvey J. Levin, "Spectrum Allocation Without Market," American Economic Review Papers and Proceedings, (May 1970), pp. 209-218; Harvey J. Levin, The Invisible Resource: Use and Regulation of the Radio Spectrum, (Baltimore: The Johns Hopkins Press, 1971).

[9] See, for example: "Note: The Crisis in Electromagnetic Frequency Spectrum Allocation: Abatement Through Market Distribution," Iowa Law Review, vol. 53 (1967), pp. 437-79; Nicholas Johnson, "Towers of Babel: The Chaos in Radio Spectrum

FCC staff members have also been writing about these ideas since 1976. [10] Similarly, other Federal government agencies and task

---

Utilization and Allocation," Law and Contemporary Problems, vol. XXXIV (Summer 1969), pp. 505-534; Arthur S. DeVany, Ross D. Eckert, Charles D. Mayers, Donald J. O'Hara and Richard C. Scott, "A Property System for Market Allocation in the Electromagnetic Spectrum: A Legal-Economic-Engineering Study," Stanford Law Review, vol 21 (1969), pp. 1499-1561; Jora R. Minasian, "Property Rights in Radiation: An Alternative Approach to Radio Frequency Allocation," Journal of Law and Economics, vol. XVIII (April 1975), pp. 221-272; Charles L. Jackson, "The Orbit-Spectrum Resource: Market Allocation of International Property," Telecommunication Policy, (September 1978), pp 179-190; Milton Mueller, "Property Rights in Radio Communication: The Key To The Reform of Telecommunications Regulation," Cato Institute Policy Analysis (June 3, 1982); Molly K. Macauley and Paul R. Portney, "Property Rights in Orbit," Regulation (July/August, 1984), pp 14-18ff. For a discussion of why the Communications Act of 1934 was written to preclude private ownership of spectrum usage rights even though such ownership rights would have been quite feasible, see: Thomas W. Hazlett, "The Rationality of U.S. Regulation of the Broadcast Spectrum," University of California at Davis, Department of Agricultural Economics, Working Paper No. 87-9 (1987).

[10] John O. Robinson, "An Investigation of Economic Factors in FCC Spectrum Management," FCC Office of Chief Engineer, Spectrum Allocations Staff, Report No. SAS 76-01, (August 1, 1976); John O. Robinson, "Spectrum Allocation and Economic Factors in FCC Spectrum Management," IEEE Transactions in Electromagnetic Compatibility, EMC-19 (August 1977), pp. 182-190; Douglas W. Webbink, "Frequency Spectrum Deregulation Alternatives," FCC Office of Plans and Policy Working Paper No. 2, (October 1980); Charles Needy, "Allocation of Scarce Resources: An Analytical Framework and Its Specific Application to Orbital Slots, Satellite Transponders and Radio Interference," FCC Common Carrier Bureau, Economics Division, Staff Report (September 1981); Rodney T. Small, "A Comparison of Alternative Spectrum Regulatory Approaches," FCC Office of Science and Technology, Spectrum Management Division, FCC/OST R82-1, (September 1982); Alex Felker and Kenneth Gordon, "A Framework For a Decentralized Radio Service," FCC Office of Plans and Policy Staff Report, (September 1983); Philip B. Gieseler, "Implementing New Technology in the Land Mobile Radio Services," FCC Office of Plans and Policy (September 1983); John O. Robinson, "Spectrum Management Policy in the United States: An Historical Account," FCC Office of Plans and Policy Working Paper No. 15 (April 1985); Evan Kwerel and Alex D. Felker, "Using Auctions to Select FCC Licensees," FCC Office of Plans and Policy Working Paper No. 16, (May 1985); James E. McNally, Jr., "Alternatives For Improving

forces have studied these issues over the years. [11] Interest in increasing the efficiency of spectrum usage has also grown for a variety of reasons such as the deregulation and privatization movements in general, and the admission by at least some licensees that in some limited situations they may have more to gain than lose by changes in the existing methods of allocating, assigning and using the spectrum. Nonetheless, many more changes remain to be made beyond those that have already taken place.

A fundamental concept of spectrum deregulation and especially the ultimate idea of establishing full private property rights in spectrum usage is that users of the spectrum should be given as much freedom as possible to use the spectrum in ways that they see fit, and to change their uses at will without seeking prior government permission for those changes as long as others are not directly harmed and as long as the public or the government is ultimately informed. Another way to describe this concept is to say that users should be given complete or at least partial property rights in the use of the spectrum. Such rights would include the authority to buy, sell, sublease, share, divide and combine their spectrum use rights; the right to exclude others from the use of the spectrum and to use the spectrum in many different ways without the necessity of first obtaining government permission. If spectrum users were given those rights they would have stronger incentives to use spectrum efficiently, i.e. to use it in ways that lead to its highest valued use.

Personal Communications," FCC Office of Plans and Policy Working Paper No. 20 (August 1986); John R. Williams, "Private Frequency Coordination In The Common Carrier Point-To-Point Microwave Service," FCC Office of Plans and Policy Working Paper No. 21, (September 1986).

[11] See for example: U.S. Department of Commerce, Commerce Technical Advisory Committee, Electromagnetic Spectrum Utilization - The Silent Crises, (1966); President's Task Force on Communications Policy, The Use and Management of the Electromagnetic Spectrum, Staff Paper No. 7 (1969); Donald R. Ewing and Leslie A. Berry, "Metrics for Spectrum-Space Usage," U. S. Department of Commerce, Office of Telecommunications, OT Report 73-24 (November 1973); Donald R. Ewing, "Controlled Markets for Spectrum Management," Proceedings of the IEEE, vol. 68 (December 1980), pp. 1536-1542. See also: Carson Agnew, Donald A. Dunn, Richard G. Gould and Robert D. Stibolt, Economic Techniques for Spectrum Management: Final Report, a study prepared for the National Telecommunications and Information Administration by Mathtech Inc. and Telecommunications Systems (December 20, 1979).

An analogy may be drawn to the ownership of land or automobiles. Although a registered deed is required to prove land ownership in most jurisdictions, and motor vehicles must be registered, have up to date license plates, be insured and the operator must possess a valid driver's license; nonetheless, land owners and vehicle owners have a very large degree of freedom in how they use their land and autos. Within broad zoning limits, land may be used for many different purposes. Within the confines of safety regulation and the prohibition of criminal activities, automobiles may be used for business or pleasure, may be used to drive between many different locations, and may be used to transport an infinite number of combinations of people and "things." [12] The ultimate goal of spectrum deregulation is to give spectrum users sufficient freedom that they will have an economic incentive to use the spectrum as efficiently as possible. The way to attain that goal would be to give spectrum licensees the same degree or at least a similar degree of property rights and hence freedom in their activities as land owners and auto owners currently possess. While this may be the ultimate goal of spectrum deregulation ideas, it is clear that neither the Commission nor the Congress have yet embraced this idea.

One point should become obvious from the description of final and proposed FCC spectrum actions that follows. The FCC has taken many deregulatory steps with respect to specific rules or specific services. But generally each decision or proposed rule change has been a separate action related to one rule or a small number of rules in one particular service or small group of services. Indeed, the pattern of FCC actions appears to be illustrative of the way in which regulatory agencies typically make changes in rules concerning the industries they regulate. The FCC has generally tended to take small marginal steps rather than make major wholesale changes in its rules at a single time. As a matter of fact, it has been argued that one of the results and perhaps even one of the intended purposes of government regulation is to slow the rate of change of industry structure and performance. [13] Thus, although the FCC has made many deregulatory changes, it is still far from allowing licensees full freedom to "let the market decide," or even from seeking the legal authority that might be necessary in order to allow the development of full private property rights in spectrum use.

---

[12] See: Douglas W. Webbink, "Radio Licenses and Frequency Spectrum Use Property Rights."

[13] Bruce M. Owen and Ronald Braeutigam, The Regulation Game: Strategic Use of the Administrative Process, (Cambridge, Mass.: Ballinger Publishing Company, 1978); and Roger G. Noll, "The Behavior of Regulatory Agencies," Review of Social Economy, vol. XXIX (March 1971), pp. 15-19.

There are many different ways in which the Commission's recent spectrum deregulatory activities could be classified. For purposes of this paper, those activities have been divided up into (A) Spectrum Use Operational Flexibility; (B) Technical Flexibility; (C) Frequency Coordination and Privately Negotiated Interference Rights; (D) Techniques for Expediting Spectrum Authorization; (E) Revenue Sources and Spectrum Pricing; and (F) Other Spectrum Use Restrictions, including License Term Length and Transferability and Petitions to Deny and Economic Impact. In a number of places, the paper includes this author's interpretation of the significance of particular deregulatory changes. The paper describes some of the legal changes that would need to be made in order to give licensees full and explicit spectrum usage property rights. Next, the paper describes why many incumbents oppose these kinds of changes. Finally, the paper includes a summary and conclusions.

II. Recent FCC Spectrum Deregulatory Activities

A. Spectrum Use Operational Flexibility

1. Provision Of Secondary Services

One of the areas in which the FCC has made significant deregulatory progress in recent years is in allowing certain services, especially broadcasting or mass media services, to provide secondary signals or services on their transmitted signal subcarrier in addition to their primary service. Generally, such services are authorized on the condition that they not harm transmission or reception of the primary service.

As an important example, when the FCC authorized FM stations to use their subcarrier or their so-called subsidiary communications authorization (SCA) not only to broadcast in stereo but also for other types of services such as paging, it greatly expanded the number of options available to such stations. As it did in many other cases, the FCC moved incrementally to authorize this service. The Commission first authorized FM stations to provide certain subsidiary services (such as news, music, time, weather and other similar programming categories) in 1955. [14] The broadcast of stereo sound was first authorized on FM stations in 1961. [15] Following that action, the Commission authorized commercial FM stations to use their subcarrier for utility load management in 1982, but at that time, load management and those services authorized in 1955 were the only allowed uses other than stereo broadcasting. [16] In 1983, the FCC expanded the allowed uses of the SCA to include an unlimited number of other broadcast, private radio and common carrier services such as paging, dispatch, data transmission and facsimile transmission. Moreover, the Commission also authorized noncommercial FM stations to provide the same kinds of services as commercial stations as long as the provision of those services did not adversely affect the provision of the radio reading service for the blind. [17]

---

[14] Subsidiary Communications Authorization, Docket No. 10832, FR 1821 (March 25, 1955).

Despite growing interest in alternative uses of the SCA, it still appears that most FM stations only provide FM stereo and do not otherwise use their SCA authority. While a few FM stations are providing a variety of other services, most are not. For example, out of a random sample of 298 FM stations in 1987, only 74 used their SCA to provide a service other than just stereo broadcasting. Of the SCA uses, the major categories were background or foreground music, data transmission and talk programs (usually the radio reading service for the blind). Interestingly, while there was, in the past, some controversy about allowing FM stations to provide paging, out of the 298 stations in the random sample, only one reported that it used its SCA to provide a paging service.

---

[15] Stereophonic Broadcasting On FM Broadcast Stations, Report and Order in Docket No. 13506, FR 3529 (April 25, 1961).

[16] Subsidiary Communications Authorization, Report and Order in BC Docket No. 81-352, 47 FR 1386 (1982).

[17] FM Subsidiary Communications Authorizations, First Report and Order in BC Docket No. 82-536, 48 FR 28445 (June 22, 1983); and Commercial Use of Public SCA's, Report and Order in BC Docket No. 82-1, 48 FR 26608 (1983).

Table 1

Number And Percent Of FM Broadcasting Stations Using A Subsidiary Carrier Authorization (SCA), 1987

	Number * of Stations	Percent of Stations in Sample
Total FM stations in sample	298	100.0%
Stations broadcasting in stereo	247	82.9
Stations not using an SCA	224	75.2
Noncommercial stations in sample	63	21.1
SCA USES: **		
Background (easy listening) or foreground (usually rock) music	30	10.1
Low, medium or high speed data transmission	20	6.7
Talk (usually radio reading service for the blind but may include instructional programs)	11	3.7
Physician's news network	4	1.3
Telemetry tones used for transmitter control	2	0.7
Digital or voice paging	1	0.3
All other SCA uses ***	12	4.0

---

* These numbers were compiled using a one in twenty (5%) random sample of data from the data source listed below.

** Note that in a few cases a station had more than one kind of SCA signal.

*** Includes relay of radio station signals, networks, cable audio, campus carrier, ethnic or foreign language, farm or financial information, gospel and religious programs, highway data, other information, sports and special events, WWV time, weather, and cases in which the use of the SCA is not indicated.

Source: Bruce F. Elving, FM Atlas and Station Directory, Eleventh Edition, (1988).

In 1980, the FCC first authorized the transmission of AM stereo. In 1982, after several attempts to choose a single AM stereo standard, the Commission decided not to mandate any single technical standard for the transmission of stereo broadcasting. [18] The decision not to mandate a specific AM stereo standard has been considered a quite controversial decision. The 1982 decision specifically allowed AM stations to transmit a stereo signal. Later in 1982 the Commission also allowed AM stations to use their subcarrier for utility load management. [19] Two years later, in 1984, the Commission adopted rules for the use of AM subcarriers that are similar to the FM subcarrier rules. AM stations are now allowed to use their subcarrier to provide a variety of services such as paging and dispatch, data transmission, etc. [20]

Similarly, in 1983 the FCC authorized the transmission of teletext in the vertical blanking interval of a TV signal. Such teletext signals are allowed to be either related (such as closed captioning) or unrelated to the main broadcast signal. [21] Noncommercial stations were also allowed to provide teletext services and sell them for a profit, but for a period of five years line 21 of the VBI was reserved for use for closed captioning for hearing-impaired persons. [22] In 1984 the use of the vertical blanking interval was expanded to include other uses such as paging, data transmission and software delivery. [23]

More recently, in 1984 the Commission authorized television stations to transmit stereo or bilingual language broadcasts on the television audio signal. While the Commission protected the pilot tone of a particular stereo system, it did not mandate any particular stereo transmission system. [24]

---

[18] AM Stereophonic Broadcasting, Report and Order in Docket No. 21313, 47 FR 13152 (March 29, 1982).

[19] Use of the AM Carrier For Utility Load Management, Report and Order in BC Docket No. 81-896, 47 FR 25342 (June 11, 1982).

[20] Use of the AM Carrier, Report and Order in BC Docket No. 83-1322, 49 FR 34011 (August 28, 1984).

[21] Teletext Transmission, Report and Order in BC Docket No. 81-741, 48 FR 27054 (June 13, 1983).

[22] Teletext Transmission, Memorandum Opinion and Order on Reconsideration in BC Docket No. 81-741, 50 FR 4678 (February 1, 1985).

[23] TV Vertical Blanking Interval Data Transmission Services, Report and Order in MM Docket No. 84-168, 50 FR 4658 (February 1, 1985).

All of these decisions illustrate cases in which the Commission authorized broadcasting stations to provide additional or secondary or subsidiary services on their main authorized channel. Generally, all of them are granted subject to conditions that they not harm the transmission or reception of the primary signal.

The interesting, and so far unresolved issue is what should be the appropriate government policy in cases in which the secondary signal might actually harm the transmission or reception of the primary signal. After all, there may be many situations in which consumers would gain and the station owner would also gain by putting more energy into the secondary signal and less into the primary signal, even if the coverage area of the primary signal decreased or the quality of the primary signal was diminished.

2. Multiple Services In The Same Frequency Block

While the examples cited above would give users substantially increased usage flexibility, they still retain the fundamental national block allocation schemes in which the primary use of a block of frequencies is for a single kind of service such as common carrier cellular radio or TV broadcasting. However, a 1985 FCC Office of Plans and Policy report suggested that the Commission should test the concept of allowing users to have even more flexibility than that described above. Under the OPP proposal, licensees could decide whether to provide a broadcast, common carrier or private radio signal within their authorized frequency. [25]

Following that OPP proposal, in 1985 the Commission proposed a so-called flexible radio service using what are presently UHF TV channels 50-59 (686-746 MHz). On those channels, according to the proposal, broadcast licensees would be free to use their channel for broadcasting, land mobile, point-to-point services or other kinds of services. Licensees, rather than the Commission, would decide the best or most profitable use of the spectrum. Essentially, licensees would be limited to causing no more interference than would a full power TV station, (or their actual 74 dBu contour in the case of a low power TV station). The Commission proposed that it be notified 30 days in advance of any change in transmitting facilities. Licensees could decide whether they wished to be treated as broadcasting, common carrier or some other kind of service. The proposal even requested comment on allowing licensees the use of channels wider than 6 MHz. [26] However, the Commission has not acted on that proposal since it issued its notice in 1985 and some parties have argued that the Commission does not have the legal authority to take such an action. [27] In addition, this is a highly controversial proposal that is opposed by many licensees, a fact that will be discussed below. In part the Commission has not acted because it is considering making supplemental VHF and UHF spectrum available for Advanced Television.

---

[24] TV Aural Broadcast Subcarrier Frequencies, Second Report and Order in BC Docket No. 21323, 49 FR 18100 (April 27, 1984).

[25] Alex Felker and Kenneth Gordon, "A Framework For A Decentralized Radio Service," FCC Office of Plans and Policy Staff Report (September 1983).

In 1986, the Commission allocated most of the remaining unallocated spectrum in the 900 MHz land mobile band. Specifically, the Commission allocated 10 MHz to the cellular service, 6 MHz to public safety land mobile, 10 MHz to private land mobile services other than public safety and authorized 2 MHz for a new general purpose mobile radio service. While the Commission concluded that it would be desirable to create a general purpose radio service which could be open to land mobile, maritime mobile and aeronautical mobile use, it did not specify assignment areas by bandwidth or area. [28] The decision indicated that the Commission would soon issue a further notice on the general purpose mobile service. [29] However, to date a further notice has not been issued.

While these proposals, if adopted, would represent substantial and important deregulatory steps, they still do not go the full distance of accepting private property spectrum rights. In particular, both proposals involve only small portions of the UHF spectrum rather than all frequencies, and both would maintain the requirement of prior government approval for changes in the kinds of signals transmitted as well as for transfer of a license.

---

[26] Further UHF TV - Land Mobile Sharing, Notice of Proposed Rule Making in Gen. Docket No. 85-172, 50 FR 25587 (June 20, 1985), pp. 25595-25598.

[27] Michael C. Rau, "Allocating Spectrum By Market: The FCC Ultra Vires?" Catholic University Law Review, vol. 37 (1987), pp. 149-189.

[28] Note: this proposal is similar to a proposal by Dunn and Owen. See: Donald A. Dunn and Bruce M. Owen, "Policy Options In Mobile Radio Spectrum Management," A Report Prepared Under Contract To The Federal Communications Commission (September 1978).

[29] 900 MHz Reserve Band, Report and Order in Gen. Docket Nos. 84-1231, 84-1233 and 84-1234, 61 RR 165 (1986); or 51 FR 37398 (October 22, 1986). That Order also held 4 MHz in reserve in the 900 MHz band and also allocated 27 MHz of spectrum in the L-band for mobile satellite service.

3. Time Sharing

Recently the demand for the Instructional Television Fixed Service (ITFS) has been growing rapidly. ITFS is a service using microwave frequencies that is designed to transmit instructional programs to colleges and universities as well as to primary and secondary schools. One reason that the demand for ITFS is growing is that in 1983 the FCC decided to allow institutions with ITFS licenses to sublet or lease time on their ITFS channels for multichannel multipoint distribution service (MMDS). The Commission also reallocated 8 ITFS channels for use by MDS. [30] MDS and then MMDS has been developed by entrepreneurs as a substitute for cable TV, i.e. as an over-the-air method for bringing movies and other entertainment channels such as Home Box Office (HBO) to consumers. While the Commission initially placed restrictions on the number of hours during which ITFS channels could be used for MMDS type purposes, [31], the FCC relaxed these requirements somewhat in 1985. [32] This has made it potentially possible to develop packages of over-the-air pay services using those channels.

In 1983, the Commission also decided to allow broadcast auxiliary facilities (such as those used for studio to transmitter links and electronic news gathering) to be used for nonbroadcast as well as broadcast related purposes. Stations were also allowed to share their broadcast auxiliary facilities. However, the Commission indicated that in the future it would no longer assign new broadcast auxiliary channels on an exclusive basis, but instead stations would be required to share those channels. [33]

---

[30] MMDS Spectrum Allocation, Report and Order in Gen. Docket No. 80-112 and 80-116, 94 FCC 1203 (1983).

[31] MMDS Spectrum Allocation, Memorandum Opinion and Order on Reconsideration in Docket No. 80-112, 49 FR 27147 (July 2, 1984).

[32] Instructional Television Fixed Service, Second Report and Order in MM Docket No. 83-523, 50 FR 26736 (June 28, 1985).

[33] Broadcast Auxiliary Facility Sharing, Report and Order in BC Docket No. 81-794, 48 FR 17081 (1983).

In 1985, the Commission authorized private operational fixed microwave systems to provide services to other eligible entities on a not-for-profit basis and to lease excess capacity on a for-profit basis. [34]

Each of these actions allowed specific groups of licensees to use their licenses for additional purposes and thus not only increased their ability to earn profits, but also increased the potential efficiency of spectrum usage.

3. Repeal of Permissible Use Restrictions

In 1985, the FCC removed restrictions on the permissible content of communications on all private land mobile channels where licensees have exclusive use of a channel. Prior to that, licensees were only allowed to communicate concerning topics directly related to the class of license for which they were licensed. However, the Commission did not repeal restrictions on permissible communications on shared channels because it feared that the level of interference might increase. Nevertheless, this was an important change in the flexibility available to private radio licensees to use their licenses as they wish. [35]

Taken together, all of these actions indicate that the Commission is gradually giving its licensees far more freedom in the operation of their facilities. Again, however, none of them has gone as far as the establishment of full explicit private property rights in spectrum use.

B. Technical Flexibility

Another way in which the FCC has substantially increased the flexibility of its spectrum regulations involves modification or repeal of specific technical regulations. Indeed, in a variety of ways, the Commission has moved to give licensees more flexibility in determining the technical characteristics of the signals they transmit. This, in turn has provided licensees with more opportunities to provide new and innovative services for which there exists or may develop in the future consumer or business demand.

---

[34] Authorizing Private Carrier Systems in the Private Operational-Fixed Microwave Radio Service, Report and Order in PR Docket No. 83-426, 50 FR 13338 (April 4, 1985).

[35] Permissible Communications Restrictions, Report and Order in PR Docket No. 84-109, 50 FR 6179 (February 5, 1985).

1. Technical Flexibility And The Question Of Mandating Technical Standards

An important and continuing debate concerns the question of under what circumstances the Commission should mandate technical standards, especially standards that involve equipment compatibility or interoperability. [36] Generally the Commission has been moving towards giving its licensees more flexibility in deciding on technical aspects of their transmitted signals. Compared to the mass media or broadcasting services, the FCC regulation of common carrier microwave and common carrier satellites has always allowed a greater amount of technical flexibility by licensees. Users have much freedom in what they can do technically. For example, when the Commission authorized the fixed satellite services, it authorized frequencies without specifying modulation type. [37] Although minimum loading standards and minimum technical efficiency rules are specified in the common carrier microwave services with respect to digital modulation, licensees are free to use a variety of modulation and channeling schemes. [38]

Gradually, the FCC has been relaxing its level of detailed regulation of technical standards in many different services. In fact, in general, the FCC has been moving towards the position that it should only mandate standards that are needed to prevent out-of-band or out-of-authorized area interference or that are necessary because of international treaty, domestic law or safety of life or property. In a 1984 Order, the Commission stated that technical regulations could be divided into four categories: interference control, spectrum efficiency, interoperability, and signal quality. [39] While the Commission indicated that control over interference was an important and continuing government regulatory function, it suggested that its regulations could be made more flexible by relying on allowing equivalent levels of interference from different forms of modulation instead of mandating specific co-channel and adjacent channel operation rules and power limits based upon conventional emission modes. With respect to spectrum efficiency, it suggested that many licensees have an incentive to be technically efficient in their use of the spectrum without the need for specific Commission regulations. Even where it is necessary that efficiency be mandated by the government, it may be desirable to specify efficiency using a measurement such as bits/second/Hertz rather than specifying particular emission modes and other detailed technical parameters. With respect to interoperability, the Commission stated that such standards can often be achieved voluntarily. However, with regard to interoperability standards mandated by law or treaty or related to safety of life or property, particularly in the maritime and aeronautical radio services, the Commission would continue to mandate rules. Finally, with respect to technical quality, in many cases the supply of services and products is sufficiently competitive that it is unnecessary for the government to mandate technical quality. Exceptions would still be made in situations involving statutory or treaty obligations or safety of life or property. In addition, in this particular 1984 Order, the Commission deleted certain specific rules involving AM modulation levels, audio frequency response, monophonic AM operations, FM performance, video signal quality, etc.

---

[36] Stanley M. Besen and Leland L. Johnson, Compatibility Standards, Competition, and Innovation in the Broadcasting Industry, (Santa Monica: The Rand Corporation: November, 1986); Yale M. Braunstein and Lawrence J. White, "Setting Technical Compatibility Standards: An Economic Analysis," The Antitrust Bulletin, vol. XXX (Summer 1985), pp. 337-355.

[37] See: 47 CFR 25.202 (1986).

[38] John R. Williams, "Private Frequency Coordination In The Common Carrier Point-To-Point Microwave Service," p. 15.

[39] Re-Examination of Technical Regulations, Report and Order in Gen. Docket No. 83-114, 99 FCC 2d 903 (1985). For a more complete discussion of these standards setting issues, see: Notice of Inquiry and Proposed Rulemaking in Gen. Docket No. 83-114, 65 FR 14339 (April 4, 1983).

a. Mass Media Technical Flexibility

When the Commission authorized AM stereo in 1982, it chose not to mandate detailed technical standards but instead decided to let market forces determine the appropriate standard or standards. [40] In part, the Commission made that decision because of the difficulty of determining and proving in court that one system was superior to all other systems. That action was highly controversial and a number of parties have claimed that the reason AM stereo has never been widely adopted was because the FCC did not choose to mandate specific AM stereo standards. [41]

---

[40] AM Stereophonic Broadcasting, Report and Order in Docket No. 21313, 47 FR 13152 (1982).

[41] Chris Sterling, "The FCC and Changing Technological Standards," Journal of Communications, (Autumn 1982), pp. 137-147; Jason B. Meyer, "The FCC and AM Stereo: A Deregulatory Breach of Duty," University of Pennsylvania Law Review, vol. 133 (December 1984), pp. 265-286.

In fact, however, it appears to this author that there has never been strong evidence that consumers desired AM stereo reception, given the relatively low audio fidelity of monophonic AM signals. Of course, since few AM stereo receivers have been manufactured, it is difficult to establish whether the lack of stations using AM stereo and the lack of receivers indicates the absence of underlying consumer demand for AM stereo, or whether the lack of consumer interest is due to the lack of stereo stations and receivers. In other words, this may be a classic case of the chicken and egg problem.

When the FCC authorized DBS in 1982, it explicitly chose not to mandate particular technical standards for DBS, other than those related to interference. [42] In its DBS order, the Commission stated that it was too early to adopt technical standards in a new untried service when there was no consensus in the industry concerning what would be the appropriate standards to choose. [43] In 1986 the Commission reiterated an earlier decision to allow non-DBS use of DBS satellites, such as data and voice communications, for a period of up to 5 years, as long as that non-DBS use did not exceed half of a 24 hour day on any DBS transponder. [44] Once again, however, the Commission chose not to mandate specific technical standards for DBS.

When the Commission authorized stereo broadcasting on TV in 1984, it decided to protect the pilot tone of a particular stereo system, the one proposed by the Electronic Industries Association and the Broadcast Television Systems Committee (BTSC) but not to mandate that system as the only approved system. This encouraged the use of the BTSC stereo system, but also appeared to allow licensees to use other systems without FCC permission, if they believed the alternative was superior. [45]

---

[42] While DBS is licensed in the Common Carrier Bureau as a common carrier service, it is expected to behave more like a Mass Media Bureau broadcast service, which is why it is mentioned in this section.

[43] Direct Broadcast Satellites, Report and Order in Gen. Docket No. 80-603, 90 FCC 2d 676 (1982). See also: Florence O. Setzer, Bruce A. Franca and Nina W. Cornell, "Staff Report On Policies For Regulation Of Direct Broadcast Satellites," FCC Office of Plans and Policy (September 1980).

[44] Permissible Uses of the Direct Broadcast Satellite, Memorandum Opinion and Order, 1 FCC Rcd 977 (1986).

[45] TV Aural Broadcast Subcarrier Frequencies, Second Report and Order in BC Docket No. 21323, 49 FR 18100 (April 27, 1984).

The Commission also indicated in 1987 that at most it might adopt nonrestrictive technical standards or even no technical standards for Advanced Television Systems (ATV). [46] In a further notice in 1988, the Commission indicated that it intended to have a role in the development of standards, but wished to allow users as much flexibility as possible to develop newer and technologically superior systems in the future. It raised the possibility of protecting some aspect of a standard but not mandating the standard, or adopting a standard only for allocation and assignment purposes, or adopting a standard with a sunset provision or encouraging the development of an "open architecture" receiver that could handle many different ATV systems. [47]

b. Private Radio and Common Carrier Mobile Radio Technical Flexibility

The Commission has also gradually relaxed many of the technical regulations concerning private radio and common carrier mobile radio services. For example, when the Commission released 250 25 KHz channels of reserved spectrum in the 800 MHz band in 1982 for use by various private land mobile services, it did not mandate any particular technical standards other than interference standards and bandwidth standards. On those channels, users are free to use any emission mode and conventional or trunked channels, as long as the 25 KHz bandwidth is maintained on single channel systems and out of band emissions are limited to the allowed limits. Licensees of multiple channel systems are allowed to subdivide channels or to use wider bandwidth channels. [48]

The Commission also recently proposed that cellular licensees be allowed to provide alternative emission modes and new spectrum efficient technologies and auxiliary services in the cellular radio service. The current cellular rules specify detailed design standards such as channel bandwidth, control channels and modulation types. Indeed, it has been argued that cellular represents a case of overregulation of technical standards. [49] However, allowing cellular licensees to use new technologies might permit them to use digital transmission that does not conform to voice characteristics, higher speed data transmission, etc. Similarly, licensees might be permitted to use time division or frequency division multiple access that is not presently permitted because of the presently required 30 KHz FM channels. [50]

---

[46] Advanced Television Systems, Notice of Inquiry in MM Docket No. 87-278, FCC 87-246, released August 20, 1987.

[47] Advanced Television Systems, Tentative Decision and Further Notice of Inquiry in MM Docket 87-268, FCC 88-288 (released September 1, 1988), para. 115-119.

[48] 800 MHz Reserve Channel Release, Private Land Mobile Radio Service, Second Report and Order in Dockets 79-191, 79-334, 79-107, and 81-703, 90 FCC 2d 1281 (1982).

In 1988, the Commission decided that it would allow licensees in the Special Mobile Radio (SMR) Service to use non-standard channel bandwidth and modulation schemes within their assigned channels. Licensees may use trunked systems or other comparably efficient methods of operations. Licensees may combine or split channels. However, loading standards will remain in effect. [51]

Earlier, in 1985 the Commission proposed a system of "Alternative Type Acceptance" for land mobile radio transmitting equipment. The current type acceptance procedure for new transmitters requires that manufacturers go through a formal FCC rule making in which the technical details of the proposed new equipment are revealed to other competing companies, before the new technology is authorized by the Commission. Hence, in the past competitors were able to copy new equipment even before it had been marketed to the public, thus decreasing the potential profits of the first entrant and consequently decreasing the economic incentive to develop new equipment. Under the newly proposed procedure, manufacturers of equipment using new technologies would not have to reveal the technical details of their equipment in advance to competitors. Instead, the manufacturer would be required to perform specific tests to compare the interference potential of the proposed equipment to existing type accepted FM land mobile transmitters. Based upon the results of those tests, the Commission would determine a "power derating factor" that would specify the maximum amount of output power the new equipment would be allowed to transmit in order to cause no more interference than existing already authorized equipment is allowed to cause. [52] In 1988 the Commission issued a further proposal on this matter. [53]

---

[49] Jeffrey Krauss, "Implications of FCC Regulation of Telecommunications Technical Standards," IEEE Communications Magazine, vol. 20 (September 1982), pp. 28-32.

[50] Liberalization of Technology in the Cellular Radio Service, Notice of Proposed Rule Making in Gen. Docket No. 87-390, 2 FCC Rcd 6244 (1987).

[51] Amendment of Part 90, Subparts M and S, Report and Order in PR Docket No. 86-404, FCC 88-69, released March 30, 1988.

c. Other Flexible Technical Standards

In 1987, the Commission proposed a revision of its regulations concerning unlicensed low power radio frequency devices operating under Part 15 of the Commission rules. One of the purposes of the proposal was to give manufacturers much more technical flexibility and to cease adopting device specific regulations, as long as radio emissions did not exceed some specified level. [54]

Taken together, these examples show that the FCC has gradually been increasing the level of technical flexibility that it allows its licensees as long as the level of interference is not substantially increased. While some of these proposals and actions have been controversial (especially AM stereo), most have been far less controversial than some of the actions discussed in the next section of this paper. Of course, it must be emphasized that technical flexibility might be desirable to many consumers even in cases where the level of interference increased significantly. [55] The Commission, however, has been reluctant to knowingly allow such situations to develop.

C. Frequency Coordination And Privately Negotiated Interference Agreements

Another important set of concepts that represent a move towards establishing private property rights in spectrum use are the use of frequency coordination mechanisms and the use of privately negotiated interference rights.

---

[52] Technical Flexibility in the Mobile Communications Services, Notice of Inquiry and Proposed Rule Making in Gen. Docket No. 85-171, 50 FR 25274 (June 18, 1985).

[53] Technical Flexibility in the Mobile Communications Services, Further Notice of Proposed Rule Making in Gen. Docket No. 85-171, FCC 88-65 (released March 22, 1988).

[54] Operation of Radio Frequency Devices Without An Individual License, Notice of Proposed Rule Making in Gen. Docket No. 87-389, FCC 87-300, released October 2, 1987.

[55] Note that many automobile drivers prefer the flexibility of driving their own auto to work even though congestion would be far lower if all drivers car pooled. That situation might be analogous to allowing increased spectrum technical flexibility, although it is by no means obvious that technical flexibility necessarily increases the overall level of interference.

1. Frequency Coordination

A spectrum deregulatory concept of great significance concerns frequency coordination by licensees. In the context of FCC regulation, frequency coordination has at least two different meanings. One meaning is that certain outside organizations in the private radio services are authorized to coordinate the use of shared frequencies so that new entrants will be licensed on the channel that has the lowest rate of utilization. However, because there are no loading limits on certain private radio frequencies, existing users generally do not have the power to prevent new entrants. Or, putting it another way, the property rights of incumbents on shared frequencies is quite limited. Thus, although the use of such coordinating organizations does lead to more equal and more intensive use of the spectrum than might otherwise occur, it cannot lead to the level of efficiency that is possible under the other type of frequency coordination.

The Commission has been using frequency coordinating organizations in the private land mobile services for many years. Generally, their job is to review license applications and recommend the least loaded frequency for applicants who use shared land mobile channels. In 1986 the Commission modified and clarified its rules concerning the use of frequency coordinating committees in the private land mobile services. Among other things, the Commission strengthened the powers and the requirements on such coordinating organizations with respect to review and approval of license applications and indicated that it would authorize only one approved coordinating organization for each private radio service such as police, fire or forest products. [56]

While this change may have increased the efficiency of the frequency coordinating process, there is also the chance that it might increase the market power of those organizations. However, the Commission stated that it would monitor the activity of these organizations for possible abuses, and would withdraw certification if necessary.

The other concept of frequency coordination, however, is the way the term is used with respect to the common carrier microwave service. There frequency coordination means that a new entrant is expected to coordinate with all existing licensees, and is allowed to enter only if the amount of interference created is acceptable to the other users, or if the new entrant pays the other existing users to accept more interference. [57]

---

[56] Frequency Coordination in the Private Land Mobile Radio Services, Report and Order in PR Docket No. 83-737, 103 FCC 2d 1093 (1986).

This frequency coordination concept could be expanded to cover many other services in which users have the right to limit or prevent the entry of additional transmitters that might cause interference to existing transmitters. This concept might be especially useful in the mass media services where it might be possible to allow additional new entry if such coordination took place.

2. Privately Negotiated Interference Agreements

The Commission has made a number of proposals to allow the idea of private interference negotiation. These ideas are very similar to the frequency coordination concept described above. There is some difference between the concepts. In common carrier microwave frequency coordination, an incumbent may completely prevent the entry of a new system in a particular location, if the new entrant is unwilling to limit sufficiently his level of interference and if he is unwilling to compensate the incumbent to accept additional interference. In contrast, privately negotiated interference rights generally are used to describe negotiations between two already licensed entities concerning interference at the edges of their coverage areas where their signals overlap. Allowing privately negotiated agreements was suggested in a 1985 FCC Office of Plans and Policy working paper that was mentioned earlier. [58] Recently, in 1987 the Commission proposed that new applicants for UHF television channels 14 or 69 could be granted construction permits only if they would protect adjacent frequency land mobile systems from objectionable interference or if they could resolve the potential interference problems through a negotiated agreement. The Commission proposed that a new UHF TV station requesting maximum authorized effective radiated power (5 megawatts) be located at least 25 miles from any land mobile facilities employing receivers operating on frequencies within 3 MHz of TV channel 14 or 69. If the criteria can not be met, the construction permit will be granted if the TV station can obtain an agreement from all existing land mobile licensees and pending applicants who would otherwise be precluded. TV stations might bear the cost to move the land mobile systems to other frequencies, adopt technical solutions to the interference problem, or compensate the land mobile users to accept interference. [59]

---

[57] Carson E. Agnew and Richard G. Gould, "Frequency Coordination and Spectrum Economics," Research in Law and Economics, vol 9 (1986), pp. 167-183; John R. Williams, "Private Frequency Coordination in the Common Carrier Point-to-Point Microwave Service," FCC Office of Plans and Policy Working Paper No. 21 (September 1986).

[58] Alex Felker and Kenneth Gordon, "A Framework for A Decentralized Radio Service."

The Commission has also proposed that it accept privately negotiated interference agreements in the AM radio band. [60] A single set of protection criteria are presently applied to all AM broadcasting stations. However, the optimal level of interference may vary from one station to another. In some cases, stations may receive interference protection in areas where no listening audience exists and the station would be willing to give up its interference protection rights. In other cases, stations may desire increased protection greater than that afforded by the general protection criteria. In either of these situations, stations might gain by the right to negotiate agreements for interference levels different than those specified in the Commission's rules.

Related to the concept of allowing privately negotiated interference agreements is the concept of improving the Commission's models and rules that predict signal coverage and expected interference. While generally the Commission does not take into account terrain shielding in protecting the coverage or interference contours of broadcasting stations and other transmitters, in one recent action the Commission decided that it would take them into account. In April, 1988 the Commission decide that in certain limited situations it would waive its rules with respect to interference caused by low power TV stations (LPTV) and translators to other LPTV stations and translators. The Commission will now consider waiver of application acceptance standards in situations in which applicants can show that interference will not be caused to others due to terrain shielding. However, it will still ignore terrain shielding effects in deciding situations in which there appear to be competing and mutually exclusive applications.

---

[59] Resolution of Interference between UHF channels 14 and 69 and Adjacent Channel Land Mobile Operations, Notice of Proposed Rule Making/Notice of Inquiry in MM Docket No. 87-465, FCC 87-336, released November 30, 1987.

[60] Review of Technical Assignment Criteria for the AM Broadcast Service, Notice of Inquiry in MM Docket No. 87-267, 2 FCC Rcd 5014 (1987).

Otherwise it will accept either assent of all potentially affected stations or a terrain study. [61]

In February, 1988 the Commission proposed to allow directional antennas by FM commercial radio stations that are short spaced from other stations compared to the normally allowed minimum distances. The proposal would allow stations more flexibility in the location of their antennas. Stations would be required to provide specified level of interference protection to stations from which they are short spaced. This proposal would allow more flexibility in the use of directional FM antennas. [62]

Recently the Commission proposed extending the top end of the AM broadcasting band from 1605 to 1705 KHz. In the existing portion of the AM band, the Commission has established a series of classes of AM stations classified by power output and antenna height. The Commission proposed that in the expanded AM band applicants would be allowed to apply to build stations in any specific location in the U.S. as long as they do not violate the Commission's one-to-a market and multiple ownership rules and as long as they do not exceed the allowed interference levels with regard to co-channel and adjacent channel stations. In its proposal for the new expanded band, the Commission also asked whether on at least some of the channels, licensees might be given a national license for use of a channel throughout the country. This would then allow the licensee to build as many or as few stations as it wished on that channel throughout the U.S.[63] Thus, in effect the licensee would be able to determine the acceptable level of interference between cochannel stations operating on its frequency.

It is interesting to note that this proposal would go in the opposite direction from previous Commission actions to allow more stations on the so-called clear channels, by reducing the potential coverage area of clear channel stations.

In July 1988 the Commission also proposed to allow certain AM stations to increase their power and thus their area of signal coverage in situations in which the stations would be subject to increased received interference but under the Commission's rules would not cause interference to other stations. [64]

---

[61] Terrain Shielding in the Evaluation of Television Translator, Booster and Low Power Applications, Policy Statement, FCC 88-160, released April 25, 1988.

[62] Short-Spaced FM Station Assignments Using Directional Antennas, Notice of Proposed Rule Making in MM Docket No. 87-121, FCC 88-73, released March 30, 1988.

[63] Planning of Broadcasting in the 1605-1705 KHz Band, Fourth Notice of Inquiry in Gen. Docket No. 84-467, FCC 88-72, released June 3, 1988.

Each of these proposals is designed to bring more reality into the prediction of expected signal coverage and interference and to allow licensees more freedom in negotiating private agreements. Hence, each one, if adopted would help licensees to adjust to local conditions without requiring the Commission to make changes in its overall national allocation and assignment rules. As such, they appear likely to enhance significantly economic efficiency.

D. Techniques For Expediting Spectrum Authorization: First Come, First Served, Lotteries, Auctions

Over the years many persons have criticized the FCC procedure of holding comparative hearings when there are two or more competing applicants for a channel or frequency. [65] Among the alternatives used or at least proposed by the FCC have been a first-come first-served approach, the use of random selection or lotteries, and ultimately the use of auctions, all as alternatives to comparative hearings. Each of these alternatives is likely to increase the speed at which licenses are granted, compared to holding a comparative hearing.

1. First Come, First Served

One of the techniques that has been widely used by the Commission in the past as a preferable alternative to comparative hearings is a first come first served system of license grants. When licenses are assigned exclusive channels but the number of available channels exceeds the number of applicants, the Commission has often used this process. This has been used in the mass media, common carrier and exclusive private radio channels. It is easy to implement such a process as long as at a particular point in time there is an excess supply of specific channels, so that all applicants can be satisfied. Of course, one way the Commission has attempted to limit the demand for channels is to impose significant financial requirements on applicants. Such requirements insure that applicants are "serious" and have the funds available to build and operate facilities, but they are also a mechanism for limiting the number of applicants. [66] For example, the Commission adopted financial qualifications for new satellite applicants in 1985, [67] and for rural cellular radio applicants in 1988. [68]

---

[64] Improving The Quality of the AM Broadcast Service, Notice of Proposed Rule Making in MM Docket No 88-376, FCC 88-252, released September 12, 1988.

[65] See footnotes 6-10 above.

One unanswered question about such financial requirements is the extent to which they raise barriers to entry to applicants who would have been able to construct and operate facilities were it not for the specific government requirements. It seems possible that such requirements do prevent entry by some individuals or firms who might have successfully operated a facility or eventually sold it to someone else who could have done so.

2. Lotteries

In 1981 Congress passed a statute which gave the Commission permissive authority to use lotteries rather than comparative hearings to select among competing applicants. [69] Among other provisions, the statute required the Commission to award preferences to certain minority groups in any lottery involving mass media services. However, the Commission initially indicated that it would be unable to use the new statute because the instructions concerning providing minority preferences for applicants in the mass media services were too vague. In addition, the statute required the Commission to determine if each applicant was qualified before it held a lottery. [70] In response to the FCC comments, the Congress passed a revised lottery statute in 1982. The second statute explicitly defined which minority groups were covered and also allowed the Commission to check qualifications only of the tentative lottery winner, rather than having to check qualifications against all lottery applicants in advance. [71]

---

[66] See, for example: Douglas W. Webbink, "Future Domestic Satellite Orbital Slot Problems: Can Frequency Spectrum Deregulation Help?" Proceedings of SATCOM '81, The International Association of Satellite Users, Annual Satellite Communications Conference, Washington, D.C., April 13-15, 1981, pp. 8-17.

[67] Licensing Space Stations in the Domestic Fixed Satellite Service, Report and Order in CC Docket No. 85-135, 50 FR 36071 (September 5, 1985).

[68] Rural Cellular Service, Fourth Report and Order in CC Docket No. 85-388, FCC 88-154, released May 18, 1988.

[69] The Omnibus Budget Reconciliation Act of 1981, Public Law No. 97-35, 95 Stat. 736-737 (1981).

[70] Using Random Selection or Lotteries Instead of Comparative Hearings, Report and Order in Gen. Docket No. 81-768, 89 FCC 2d 257 (1982).

The Commission held its first low power TV lotteries and SMRS lotteries in 1983, [72] its first cellular radio service lotteries in 1984, its first MMDS lotteries and private radio operational fixed service (OFS) microwave lotteries in 1985, and its first digital termination service (DTS) and digital electronic message service (DEMS) lotteries in 1986. It has been holding many lotteries since then. However, the Commission has not held any lotteries for licenses in a number of major communications services such as communications satellites and full power radio and TV broadcasting stations. Because it is relatively easy and low cost to enter a license lottery as opposed to a comparative hearing, many applicants have been entering those lotteries, and the number of applicants in successive lotteries has risen.

Table 2 shows the increasing number of applicants per lottery held in the low power TV service. Table 3 provides similar data for common carrier cellular radio service lotteries. Table 3 is particularly striking because it illustrates the fact that generally the number of applicants per lottery has grown rapidly over time, even though the newer applicants are applying in smaller and smaller metropolitan areas where the possibilities of obtaining large numbers of subscribers and hence large profits generally get smaller. [73] Both of these tables illustrate the fact that more and more people are realizing that radio licenses won in an FCC lottery are a valuable and easily resalable right.

---

[71] The Communications Amendments Act of 1982, Public Law No. 97-259, 96 Stat. 1092 (1982).

[72] Using Random Selection Instead of Comparative Hearings, Second Report and Order in Gen. Docket No. 81-768, 93 FCC 2d 952 (1983).

[73] Recently the Commission received an average of 525 applications for each of 5 different rural service areas in Nevada, an average of 557 applications for each of 6 different rural service areas in Utah, and an average of 563 applications for each of 5 different rural service areas in Wyoming. See: "Common Carrier Public Mobile Services Information, RSA Application Receipts," FCC Public Notice, no. 4292 (August 25, 1988).

Table 2

Low Power TV License Application Lotteries, 1985-1987

Fiscal Year	Number of Separate Lotteries Held	Number of Mutually Exclusive Applications	Average Number of Applications Per Lottery
1985	438	3,281	7
1986	572	6,275	11
1987	434	7,522	17

Source: FCC, Annual Report, various years.

Table 3

Number of Cellular License Lottery Applications, 1982-1986

Market Rank	Filing Date	Total Number of Applications	Average Number of Applications Per Market
1- 30	6/82	194	6
31- 60	11/82	396	13
61- 90	3/83	567	19
91-120	7/84	5,178	173
121-135	2/86	8,007	534
136-150	2/86	7,436	496
151-165	3/86	6,367	424
166-180	4/86	8,471	565
181-240	5/86	24,018	400
241-305	5/86	37,650	579
		99,284	

Source: Brian Deevy, "The Cellular Lottery Has Cost Consumers and the Industry Dearly," Radio Communications Report, (September 15, 1986), p. 34.

With so many persons applying for licenses in license lotteries, it is not surprising that there have been a number of fraudulent schemes involving convincing people to pay others to help them enter lotteries. [74] Members of Congress and industry participants are disturbed by the large number of applicants, especially applicants who apparently have no intention of building a facility, but only reselling their license for a profit.

Nevertheless, the large and growing number of applicants merely illustrates dramatically the fact that the FCC is giving away valuable spectrum use rights almost for free. Despite the large number of applicants, lotteries still seem a far faster way than comparative hearings to get licenses to the public. Moreover, the fact that many licenses are resold merely means that the initial winner, rather than the government, is holding an auction in spectrum use rights and the initial winner, not the government obtains any windfall profits.

3. Auctions

The current and the previous Chairman of the Federal Communications Commission as well as a few members of Congress have proposed the use of auctions for new licenses for non-broadcasting uses of the spectrum. [75] Despite the many desirable attributes of auctions, there appears to be little likelihood that Congress will grant the Commission legislative authority to conduct auctions in the near future.

Using auctions to award licenses or spectrum use rights would be desirable for a variety of reasons. Auctions are far more likely than comparative hearings to lead rapidly to efficient use of the spectrum. Because the winner is generally the highest bidder, the winner is likely to be the bidder who values the use right the highest and thus is likely to be the user best able to provide the highest valued service to consumers or businessmen. Since applicants know they will have to be the high bidder in order to win, the number of applicants tends to be self limited to those willing to pay the most. Hence, unlike lotteries, there are likely to be only a small number of applicants in a spectrum auction. Hence auctions are likely to get service to the public faster than lotteries and far faster than comparative hearings. In addition, of course, auctions provide a source of revenue to the government. [76]

---

[74] See, for example: Fleming Meeks, "Cellular Suckers," Forbes (July 25, 1988), p. 41.

[75] See: Extension of Remarks by the Hon. Don Ritter on the "Spectrum Assignment Act of 1988," Congressional Record (August 4, 1988), pp. E2617-E2618. See also: Letter from Dennis R. Patrick, Chairman, FCC, to the Honorable Daniel K. Inonye, Chairman, Subcommittee on Communications, U.S. Senate, April 28, 1987; and Statement of Mark S. Fowler, Chairman, FCC, before the Subcommittee on Telecommunications, Consumer Protection and Finance, Committee on Energy and Commerce, U.S. House of Representatives, October 1, 1986.

However, as mentioned above, there seems to be little or no chance that in the near future Congress will pass a statute to explicitly authorize auctions. Nevertheless, it is worth noting that it has been argued that the Commission presently has the statutory authority to use auctions to select licensees from among license applicants. [77] On the other hand, no one seems to have argued that the Commission has the legal authority to auction off permanently spectrum usage property rights. Thus, even though the use of spectrum auctions seems unlikely in the near future, one can only hope that eventually Congress will give the Commission the required statutory authority.

Each of these alternatives: first come, first served, lotteries and auctions provide the opportunity to make radio licenses available to the public far faster than comparative hearings. While auctions seem to be particularly efficient, lotteries can also allow licenses to be issued quickly. However, the efficiency of spectrum usage would be increased even more if the Commission would allow lottery winners to resell immediately their license. While the Commission has generally opposed such immediate resale following a lottery because it makes it obvious that people are "speculating" on winning the lottery, in fact, that _is_ what is happening anyway. Imposing a waiting period for resale may just slow down the rate at which spectrum usage is transferred to those who value it the most because they will use it in its highest valued use.

---

[76] Evan Kwerel and Alex D. Felker, "Using Auctions to Select FCC Licensees," FCC Office of Plans and Policy Working Paper No. 16 (May 1985); Carson E. Agnew, "Alternative Radio Licensing Arrangements: A Case Study of the Multipoint Distribution Service," _Telecommunications Policy_, (June 1983), pp. 121-136; John O. Robinson, "Assignment of Radio Channels in the Multipoint Distribution Service by Auction," _Proceedings of the Sixth Annual Telecommunications Policy Research Conference_, Herbert S. Dordick, editor (Lexington: Lexington Books, D.C. Heath, Inc., 1979), pp. 379-391.

[77] Reply Comments of Henry Geller and Donna Lampert in Report No. DS-265, May 23, 1984; and Reply Comments of Henry Geller and Donna Lampert in CC Docket 83-1096, December 30, 1983.

E. Revenue Sources and Spectrum Pricing: License Fees, Spectrum Fees And Auctions

1. Cost Of Service License Fees

In 1976, a series of Court of Appeals decisions overturned a previous FCC fee schedule. [78] Eventually the FCC returned much of the license fee money that it had collected over a number of previous years. [79]

In 1985, at the request of the Commission, Congress passed a statute that explicitly empowered the Commission to collect cost of license processing and service authorization fees and specified the level of fees for various specific services. The fees were intended only to cover a portion of the administrative cost of granting various kinds of licenses and other Commission services. [80] In April 1987, the FCC instituted the new fees. [81] The fees range from $30 for a land mobile radio license to $200 for a cellular system construction permit to $2,250 for a commercial TV station license to $18,000 for an application to launch and operate a communications satellite. The Commission is also authorized to review the fees every two years to take into account the effect of inflation as reflected by changes in the Consumer Price Index for All Urban Consumers. In the first full year of the license fee program, the Commission collected about $21,700,000 in license fees. Revenue collected from the fee program amounted to about 22% of the Commission's FY 1987 budget appropriation of $97,099,000. [82]

As was indicated above, it is important to understand that these fees are only intended to cover a fraction of the out-of-pocket costs of FCC license processing, and are not intended to act as a "spectrum fee" or to take into account externalities or other costs imposed or benefits conferred on those who use the spectrum.

---

[78] National Cable Television Association v. FCC, 554 F. 2d 1109 (D.C. Cir. 1976); National Association of Broadcasters v. FCC, 554 F. 2d 118 (D.C. Cir. 1976); Capital Cities Communications v. FCC, 554 F. 2d 1135 (D.C. Cir. 1976).

[79] See: Fee Refunds and Future FCC Fees, Notice of Inquiry in Gen. Docket No. 78-316, 43 FR 46658 (1978); First Report and Order in Gen. Docket No. 78-316, 44 FR 21267 (1979); and Second Report and Order in Gen. Docket No. 78-316, FCC 80-496 (released April 29, 1980).

[80] The Consolidated Omnibus Budget Reconciliation Act of 1985, Public Law No. 99-272, 100 Stat. 82 (1986).

[81] Fee Collection Program, Report and Order in Gen. Docket No. 86-285, 52 FR 5285 (February 20, 1987).

[82] FCC, Annual Report, Fiscal Year 1987.

2. Spectrum Fees

One issue that has been raised on occasion is whether it would be desirable to charge license fees which, rather than merely covering the cost of processing of licenses, actually acted as a spectrum fee and were related to the value of the spectrum used. Despite the imposition of modest license fees, the number of competing lottery applicants for certain desirable licenses such as cellular radio has continued to increase rapidly over time. One obvious idea, if auctions are not considered feasible, would be to raise the level of license fees substantially. Thus, for example, the fee to enter a cellular lottery in large markets could have been raised from $200 to $2,000 or $20,000 or even $200,000. However, such a large adjustment would obviously require a change in the fee statute, in order to allow the FCC to adjust spectrum fees to reflect the value of the spectrum. Such fees would then be charged according to the value of the spectrum in alternative uses.

In other contexts, Federal, state and local governments user fees are charged for the provision of many different services. [83] While some of those fees are intended mainly to recover paper processing costs, others may take into account cost concepts such as externalities, opportunity cost of providing the service, etc. It is possible to imagine taking into account similar factors with a spectrum fee. However, one real problem with establishing such fees is that it is difficult to figure out how one would accurately calculate such fees, and how one would sustain the fees actually charged when they were challenged in court. [84]

Senator Hollings and Inouye recently proposed a bill, S. 1935, that would impose a 2% to 4% tax on the transfer of all station licenses. [85] While the bill would propose a 2% tax on all ordinary license transfers, 4% would be charged on broadcasting stations transferred within 3 years of a previous sale. An additional 1% would be assessed on any broadcasting station found to have violated the Fairness Doctrine. A trust fund would be established to fund public broadcasting and all moneys collected would go into that fund.

---

[83] Clayton P. Gillette and Thomas D. Hopkins, "Federal User Fees: A Legal and Economic Analysis," *Boston University Law Review*, vol. 67 (November 1987), pp. 795-874. See also: *User Fees and Charges For Public Lands and Resources*, A Study Prepared by the Staff of the Public Land Law Review Committee (December 1970).

[84] See: Douglas W. Webbink, "The Value of the Frequency Allocated to Specific Uses," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-19, (August 1977), pp. 343-351.

Such a system of fees would therefore be based on the sales value of the license plus all physical assets rather than the value of the spectrum itself. Hence, such a fee does not actually constitute a spectrum fee, but is more similar to an excise tax, since it is unrelated to the value of the spectrum. Or putting it another way, while this proposal might be a way to raise additional tax revenue, it would not necessarily have any positive impact on the efficiency of spectrum usage.

3. Auctions

Auctions have repeatedly been suggested as an alternative to comparative hearings, first come first served, lotteries or license fees. Although the principle reason for encouraging auctions would be to increase the speed of assignment and licensing and increase the economic efficiency of spectrum use, spectrum auctions could also be a source of revenue to the government. That the auction of the spectrum might be a source of substantial revenue is suggested by the experience with the auction of other government owned or controlled resources. For example, large amounts have been collected in offshore oil and gas lease auctions in recent years. In the last few years, the amount paid in winning bids in offshore oil and gas lease auctions ranged from a low of $187 million in 1986 to a high of $6.653 billion in 1981. [86] Auctions have also been proposed repeatedly as a way to sell landing slots for airlines in congested airports. [87] In a 1986 letter to Congress, FCC

---

[85] Broadcasting, (May 2, 1988), pp. 68-69.

[86] These figures only include the bonus bids for offshore oil and gas lease rights and do not include annual rental or royalty payments on any oil or gas that was subsequently recovered. See: U.S. Department of the Interior, Minerals Management Service, Offshore Information and Publications, Federal Offshore Statistics: 1986, OCS Report MMS 88-0010 (1988), table 16, p. 44.

[87] Robert M. Hardeway, "The FAA 'Buy-Sell' Slot Rule: Airline Deregulation At The Crossroads," Journal of Air Law and Commerce, vol. 52 (Fall 1986), pp. 1-75. See also: Donald Koran and Jonathan D. Ogur, Airport Access Problems: Lessons Learned From Slot Regulation By The FAA: An Economic Analysis, Federal Trade

Chairman Fowler suggested that auctions of unused land mobile spectrum licenses to be awarded in the next two years might yield between $343 million and $1.9 billion. [88]

While the Commission began collecting cost of service fees in 1987, these fees cover only a small fraction of Commission expenses. They do not begin to reflect the value of the spectrum used by licensees. Auction of spectrum rights or spectrum fees that more closely reflected spectrum value could easily raise hundreds of millions of dollars if not billions of dollars per year in revenues. It is easy to believe that licensees would be willing to pay somewhat higher fees for faster and generally improved service from the FCC. More importantly, higher fees or auctions might act as a far more efficient rationing device than comparative hearings and Commission allocation proceedings.

F. Other Restrictions On Spectrum Usage

A variety of other kinds of past restrictions have limited the value of spectrum usage permits, i.e. radio licenses. One class of those restrictions involve the length of licenses and restriction on transfers of licenses. Another class involves the use of petitions to deny in anticompetitive ways to prevent or slow the renewal or transfer of licenses.

1. License Term Length And Transferability

In 1982 the Communications Act was amended to allow the Commission to extend the length of nonbroadcast radio licenses from 5 to 10 years. The law also increased the length of license term for radio broadcast stations from 5 to 7 years, but left the maximum length for television broadcast stations at 5 years. [89] In response to the law, the FCC increased the term of most communications licenses to the maximum length allowed in the statute.

This change had the result of increasing the time period during which most licensees could be certain of retaining their license and thus increased the value of their license, despite the fact that the Communications Act indicates that no one may assert ownership rights in the spectrum. Of course, it is true

---

Commission, Bureau of Economics (May 1983).

[88] Letter from Mark S. Fowler to the Honorable Timothy E. Wirth, dated November 21, 1986.

[89] Communications Amendments Act of 1982, Public Law No. 97-259, 96 Stat. 1093 (1982).

that the Commission rarely fails to renew licenses, but nevertheless, the threat exists. Hence, these changes in the length of license terms are likely to increase the efficiency of system usage by allowing licensees to engage in more long term planning and investment.

In 1982 the Commission repealed the 3 year broadcasting antitrafficking rule and substituted a 1 year rule for licenses obtained in a comparative hearing or a license lottery. [90] However, licenses obtained by other means, such as those purchased from another licensee, may be resold at will (following Commission approval). In 1983 the Commission also repealed its antitrafficing rules with respect to common carrier paging systems. [91]

These modifications of the antitrafficing rules were desirable because they made it easier for a licensee to sell his or her license and get out of the communications business if some one else valued the license more. That possibility reduced the risk of entering the communications business in the first place and thus is likely to increase the number of potential entrants and the level of competition in the industry.

In addition to modifying its antitrafficing rules, in 1982 the Commission authorized the sale of certain satellite transponders without requiring the owner to sell the entire satellite. [92] The purchase of a transponder is an efficient alternative to a firm individually owning and operating its own communications satellite with an FCC license or leasing a transponder from a satellite owner under a common carrier tariff.

Thus, this represents another situation in which one entity is, in reality, selling its (admittedly limited) property right to another entity. This Commission decision lowered the risk to individual companies who desired to launch satellites and thus increased their potential profits, since sale of transponders is now an alternative to leasing. However, the Commission retained it ability to review and approve or disapprove such sales.

---

[90] Broadcast Station Voluntary Assignments or Transfers of Control, Report and Order in BC Docket No. 81-897, 47 FR 55924 (December 14, 1982).

[91] Public Mobile Services Rule Revision, Report and Order in CC Docket No. 80-57, 54 RR 2d 1661 (1983).

[92] Domestic Fixed Satellite Transponder Sales, Memorandum Opinion, Order and Authorization, in CC Docket No. 82-45, 90 FCC 2d 1238 (1982).

2. Petitions To Deny And Economic Impact: The Carroll Doctrine And UHF Impact Statements

In 1987 the Commission repealed the Carroll Doctrine and the UHF impact policy. [93] Under the Carroll Doctrine, the Commission could receive petitions to deny the granting of a new license authorization by existing licensees asking the Commission to consider the adverse economic impact on them of authorizing a new competing station in their market. However, the Commission has never actually denied a license application based upon a Carroll Doctrine finding of adverse economic impact. Under the UHF impact policy, the Commission could consider the adverse impact on UHF stations of authorizing new or improved VHF service. In its decision, the Commission found that UHF service has improved to a point at which such protection for UHF stations was no longer needed. Moreover, the Commission found that both policies disserved the public, i.e. viewers, by delaying the provision of new television service. In addition, the Commission indicated that it was prepared to rely more on market forces to determine the profitability of broadcasting stations. However, the decision makes it clear that it is only applicable to applicants for full service stations and does not cover translators or low power TV stations.

Nevertheless, this represents a highly desirable change because at least in this case the Commission indicated that the desires of consumers rather than the wishes of incumbent licensees were key in its decision making. It also represents the removal of one set of potential restrictions on the entry of new spectrum users.

III. Establishing Full Property Rights and the Need to Revise The Communications Act of 1934

While the changes made and the proposals before the Commission indicate that it has taken many deregulatory and pro-market steps, there is still a great deal more that could be done.

One ultimate goal of those who favor deregulation of spectrum allocation, assignment and use (including the author of this paper) would be to grant complete or nearly complete private property rights to radio licensees. In essence, under such a proposal, a licensee would be given the same kinds of property rights as a land owner or the owner of an automobile who also possessed a valid driver's license. Such owners have the right to buy, sell, rent, lease, share, subdivide and combine their property rights. They could also exclude others from their property. They could also use it for a wide variety of legal purposes without obtaining prior government approval. Of course, it is true that land use is often subject to zoning restrictions. Nevertheless, the idea would be that in essence, the distinctions between spectrum allocated for mass media v. common carrier v. private radio uses would disappear. In the case of radio and TV broadcasting stations, an additional and related goal would also be to give licensees the same degree of First Amendment rights as those currently given to newspapers and magazines. [94]

---

[93] Policies Regarding Detrimental Effects of Proposed New Broadcasting Stations on Existing Stations, Report and Order in MM Docket No. 87-68, 3 FCC Rcd 638 (1988).

To allow such property rights to exist legally, even though they may now at least partially exist de facto, it appears that a substantial number of sections of the Communications Act of 1934 would need to be modified or repealed. [95] In several places in the Act, including sections 301, 304 and 309(h), the Act explicitly prohibits private ownership of the spectrum. These sections would need to be repealed.

To give users full private property rights in spectrum use, the right of other parties to file petitions to deny against initial licenses or license renewals under section 309 would also need to be repealed. There is no logical reason why the Communications Act should give persons the right to file such petitions against station applicants or owners when individuals have no similar right to file petitions against auto owners or land owners. [96] Certainly the typical automobile is more likely to cause physical harm to individuals than can the typical broadcasting station. While it may be true and has certainly been argued that broadcasting stations may sway public opinion, newspapers obviously do the same thing and there is no petition to deny process than can be invoked against newspapers.

---

[94] Mark S. Fowler and Daniel L. Brenner, "A Marketplace Approach To Broadcast Regulation," Texas Law Review, vol. 60 (February 1982), pp. 1-51.

[95] For one review that argues that even to allow licensees flexibility to provide TV or land mobile service on channels 50-59 of the UHF system, the Communications Act would need to be revised, see: Michael C. Rau, "Allocating Spectrum By Market Forces: The FCC Ultra Vires?" Catholic University Law Review, vol. 37 (1987), pp. 149-189.

[96] Admittedly, it is true that individuals use the municipal planning and zoning process to slow or block the authorization of new uses of land, but normally that process can not be used to block or slow land or home transfers or purchase for already approved uses.

If, as is implied above, radio licenses were to be granted with no more difficulty than a driver's license, then all references to the requirement that the Commission make a finding of "public interest, convenience and necessity" before making an initial license grant or a license transfer that appear in Sections 301, 307, 309(a) and 310(d) of the Communications Act would need to be repealed. The ultimate idea would be that there would remain no prior government review of license transfers, just as there is no prior review of transfers of ownership of automobiles or homes. At most, there would remain a requirement that an after the fact information filing be made with some government agency such as the FCC.

Similarly, Section 307(b) of the Communications Act contains the provision concerning the distribution of licenses among the several states which has been taken to encourage localism through cross subsidies for small communities, especially in the mass media services. While this provision may indeed cause the construction of some stations in more rural or less populated areas than would otherwise take place, there is little evidence that it reflects the desires of consumers. Consumers probably prefer mass entertainment programs to locally produced programs in many cases (with the possible exception of some local news shows). Hence, in a world of full spectrum property rights, this provision should also be repealed.

In addition, it would be desirable to repeal the provision in section 310(a) that prohibits foreign ownership of radio stations, even though repeal of that section is not essential to establishing private property rights in spectrum. Just as there is no limit on foreign ownership of land or autos there is no convincing reason why there should be a limit on foreign ownership of stations, and much to be gained by repealing such a rule. If only a few competing stations or transmitters existed, there might be some reason to be concerned about foreign ownership of broadcasting stations. One reason might be that stations would be used to sway public opinion to engage in illegal or treasonous acts. But there are other criminal statutes available to deal with such actions. But in a world of many stations, there is no obvious reason to retain this concern. Moreover, section 706 of the Communications Act is available to give the President the power to deal with national emergencies such as a war.

This brief list of recommended changes in the Communications Act of 1934 is not intended to be all inclusive. Nevertheless, it does include many of the provisions of the Act that are most constraining in terms of legally recognizing spectrum property rights.

IV. Why Incumbent Licensees Often Oppose Spectrum Deregulation And Spectrum Markets

An interesting issue is why so many incumbent licensees oppose many of these spectrum deregulation and spectrum property rights or spectrum market ideas. Obviously those licensees believe that they would be worse off in a deregulated market oriented system than in the present regulatory system. [97] The question is: Why?

It certainly appears that the present system restricts or slows the entry of new competitors and competing services compared to a freer market oriented system. Naturally, one should not expect incumbents to desire more competing firms in their industry, since when those new entering firms might dissipate economic rents presently being earned by incumbents. Indeed, incumbents are often particularly good at playing "the regulation game," [98] as a technique to restrict potential competitors.

Of course, there is also the equity issue that persons who bought licenses from existing licensees may have already paid for those rents in the capitalized value of the license. Hence those individuals might suffer a substantial capital loss if deregulation allowed and encouraged many new entrants that bid away the existing economic rents. [99]

---

[97] See, for example: "How Heavy A Hand Should The FCC Wield," Broadcasting (September 12, 1988), pp. 70-72; National Association of Broadcasters, "Is More Necessarily Better: White Paper," (June 1988); Letter from the Association of Broadcast Engineering Standards, Association of Independent Television Stations, Inc., AMST, CPB, NAB and NPR to Chairman Dennis R. Patrick, dated August 3, 1988; and "Broadcasters Call For Spectrum Sanctity," Broadcasting, (August 8, 1988), p. 24. See also: "LMR Groups Unite to Fight Flexible Allocation Proposal," Radio Communications Report (April 15, 1986), p. 1ff; and "Flexible Allocation Denounced by Senate Group," Radio Communications Report (June 15, 1986), p. 1ff.

[98] See: Owen and Braetigam, The Regulation Game.

[99] A key issue is the extent to which a regulatory agency should be concerned with equity rather than efficiency. See: Arthur Okun, Equality and Efficiency: The Big Tradeoff (Washington, D.C.: The Brookings Institution, 1975).

Obviously, an implicit assumption in most of this paper has been that an agency's principal goal should be to increase economic efficiency. For a rather different view, see: William H. Melody and Dallas W. Smith, "Radio Spectrum Allocation: Role

Similarly, users fear that lotteries and especially auctions may illustrate the obviously high value of rights being given away by the government through license grants. Existing licensees oppose auctions, even if they would only apply to new applicants. If the American public fully realized that some radio licenses are worth millions of dollars when initially given out, the public might press to have all commercial licensees pay substantially more for their right to use the spectrum. We should not expect anyone to volunteer to pay for a right or a service that they now appear to receive "for free." And yet, since even the military and police and fire departments are required to pay for land, buildings, automobiles, computers, etc., there is no obvious public policy reason not to auction off spectrum rights or explicitly charge users the opportunity cost for the use of the spectrum.

Perhaps what is more surprising is that many incumbents oppose being given more technical and operational flexibility in how they may use their licenses. At first glance, it is hard to understand how any business person could oppose a change which gives him or her more options and flexibility. Such opposition must come from the fact that each user group fears the information that such action might provide to others and the possible response by Congress or the FCC. For example, if broadcasters were given the freedom and then chose to switch their transmitters to land mobile usage, that might imply to the FCC that land mobile use is more valuable than broadcasting, and that might in turn lead to the Commission to reallocate spectrum away from broadcasting and towards land mobile. Perhaps broadcasters would not be so wary of flexible use ideas if they were not also afraid that the FCC might later completely reallocate spectrum uses.

In addition, licensees have a real and, at least in some instances, legitimate fear that new services, technical deregulation and new entry might increase the level of congestion and interference on the bands, especially on the AM broadcast band. Of course, if licensees could be fully compensated for any increased congestion and interference, they might be perfectly willing to accept it, but in many of these proposed changes there is no established mechanism for compensating incumbents who might lose something of value. On a number of occasions it has been pointed out that in any deregulatory activity, an important

---

of the Market," American Economic Review Papers and Proceedings, 70 (May 1980), pp. 393-397; and William H. Melody and Dallas W. Smith, "Opportunity Cost and Radio Spectrum Allocation, A Study of Feasibility of Applying The Opportunity Cost Concept To The Spectrum Allocation Process," prepared for the Department of Communications, Ottawa, Canada, Contract No. OSU77-00368, (March 1978).

requirement to obtain support is to establish a mechanism to compensate losers. [100] This issue has received little attention by persons proposing spectrum deregulation.

Another issue is that a wide variety of past Commission activities were intended to force broadcasting stations to cross subsidize certain kinds of programs such as news or public affairs by requiring stations to present such programs whether or not stations wished to do so and whether or not consumers desired those programs. The radio deregulation and television deregulation proceedings cut back on those requirements. [101] Nevertheless, some so-called "public trustee" requirements still exist. For example, there are certain provisions of the Communications Act such as sections 307(b), and 319 that were explicitly designed to give subsidies to certain persons or groups or areas of the country. Obviously those individuals will resist any attempt to take away their subsidy. In addition, to the extent that increasing flexibility makes it easier for more entities to enter the industry and to compete, it will be increasingly difficult for broadcasting stations to cross subsidize such programming or services. Incumbents who believe that such mandated cross subsidies are desirable may object to such new entry, but other incumbents will use such "merit goods" arguments as a way to maintain a purely protectionistic and anticompetitive goal of limiting competition.

Another issue may be that some firms do not like to share the spectrum with other firms, even if there would be substantial cost savings by doing so. For example, public safety organizations (e.g. police departments) prefer exclusive assignment of frequencies and exclusive control over those frequencies, and especially prefer not to share channels with non-public safety users. Police departments evidently feel that when they need to use a channel in a crisis situation, they do not want to risk that there is another, possibly interfering user, on the channel. Of course, there are technical solutions for taking over control of a channel in an emergency situation, even if the channel is shared in nonemergency periods. Moreover, the increasing use of cellular radios by police departments suggests that at the same time they do desire access to flexible systems that would allow them to be able to contact many different and unrelated parties.

---

[100] Kenneth Gordon, "Deregulation, Rights, and the Compensation of Losers," in Economic Regulation: Essays in Honor of James R. Nelson, edited by Kenneth D. Boyer and William G. Shepherd, MSU Public Utilities Papers (East Lansing: Michigan State University, 1981), pp. 223-244; Darius W. Gaskins, Jr. and James M. Voytko, "Managing The Transition To Deregulation," Law and Contemporary Problems, vol. 9 (Winter 1981), pp. 9-32; Alfred K. Kahn, "Applications of Economics to An Imperfect World," American Economic Review Papers and Proceedings, vol 69 (May 1979), pp. 1-13.

[101] Deregulation of Radio, Report and Order in BC Docket No. 79-219, 46 FR 13888 (February 24, 1981); and Television Deregulation, Report and Order in MM Docket No. 83-670, 49 FR 33588 (August 23, 1984).

In addition, as one mass media attorney suggested to me years ago, broadcasters (at least) can deal with almost any set of government regulations, whether they like them or not. What incumbent broadcasters often find most difficult to deal with are rules that are constantly changing, so that licensees never know precisely what are the rules of the game. It may well be that this fear of uncertainty and change explains why some licensees even oppose ideas such as the authorization of secondary services (such as alternative uses of AM and FM subcarriers) that would appear to provide them with added benefits but no increased costs.

IV. Conclusions

Although the FCC has made many deregulatory changes and proposed many more that have not yet been enacted, there is still a very long way to go if spectrum users are to be given full freedom to use the spectrum in the ways which they wish and which they believe would be most profitable. There is even a longer way to go before the Congress passes legislation that allows full spectrum property rights to be explicitly recognized. As this paper has indicated, the Commission has made major deregulatory strides with respect to many of its technical regulations and standards. Increasingly the Commission is allowing entities to provide secondary services on their primary signal and to have flexibility in the kinds of emission modes they transmit. However, the Commission has remained far more reluctant to break down the block allocation barriers. For that reason, the concept of flexible usage of the spectrum remains very controversial. Similarly, the concept of allowing users to negotiate interference rights different than those initially granted by the FCC, particularly in the broadcast band, remains controversial, even though the Commission could retain jurisdiction to review and approve or disapprove all such agreements.

One of the most promising market based ideas is the increased use of frequency coordination mechanisms. This has been successfully used in the common carrier microwave and satellite services and could be expanded to other services using existing Commission authority under the Communications Act. This idea is very similar to the concept of privately negotiated interference agreements.

While the Commission has lengthened license terms (in response to a new statute) and repealed the Carroll Doctrine and UHF impact statement, there still remains concern with the economic viability of licensees. In a full spectrum flexibility and deregulatory world, the economic viability of any individual firm or even any group of firms would not be of concern to a government agency.

While the Commission has made wide use of lotteries to authorize new service and has instituted modest license fees (in response to receiving explicit statutory instructions), initial licensees still do not pay anything close to the full opportunity cost of the spectrum in congested urban areas. Of course, licensees who purchase a license from someone else do pay much closer to the opportunity cost of the spectrum, taking into account the fact that there remain many restrictions on the use of the spectrum that limit the value of the spectrum in alternative uses and thus limit the estimation of its opportunity cost. Nevertheless, because many licensees do not explicitly pay what the spectrum is worth in terms of its best alternative use, many licensees do not have an incentive to use the spectrum as efficiently as they otherwise might. The lack of an explicit price on the spectrum and the resulting inefficient use of the spectrum suggests that there would be great benefits to raising the level of license fees, and even more benefit to instituting auctions for spectrum usage rights. It is unclear when or if Congress will explicitly authorize the Commission to use auctions or if the Commission will ever attempt to use auctions without explicit statutory authority.

Licensees still do not have full rights to subdivide and combine blocks of spectrum, to share, lease and sell spectrum at will and to use any technology and provide any service using spectrum as long as they do not explicitly harm (i.e. interfere with) other users. As a result, licensees' incentives to develop new services and use new technologies is muted compared to what it could be in a world of real explicitly recognized private property rights. Developing true private property rights and unrestricted spectrum markets is still a long way off. Nevertheless, many desirable deregulatory actions have been taken, and many more have been proposed. [102]

---

[102] The U.S. is not the only country that is looking at the possibility of further spectrum deregulation. See, for example: CSP International, Deregulation of The Radio Spectrum In The UK, Commissioned by the Department of Trade and Industry (London: Her Majesty's Stationary Office, 1987).

ATV STANDARDS AND TRADE IN RECORDED VIDEO ENTERTAINMENT

Steven S. Wildman
Department of Communication Studies
Northwestern University

Prepared for

The Sixteenth Annual Telecommunications
Policy Research Conference

Airlie House
Airlie, Virginia
October 30-November 1, 1988

I. Introduction

It has been recognized for some time that choices among technical standards may affect significantly the outcomes of market competition. For this reason the activities of private, voluntary standard setting organizations have been held subject to review by U.S. antitrust authorities.[1] Given the dramatic decline of the U.S. consumer electronics industry in the face of foreign competition and current levels of concern over persistently high trade deficits, it is not surprising that various commentators have advocated selecting standards for advanced television (ATV)[2] that are best suited to the competitive interests of the U.S. firms that would manufacture the hardware necessary to produce, transmit, and receive advanced television programs . It will cost billions to replace the stock of NTSC equipment currently employed by consumers and firms.[3] Some see a chance for the revival of the United States consumer electronics industry in the investment that will be required by ATV.

Virtually ignored in the discussion of trade strategies is the possibility that the standards that eventually emerge may also affect trade in television programs and films–the software that is produced, transmitted, and displayed by the various types of hardware that are the primary focus of concern. While it has received little attention, the consequences of ATV standards choices for trade in films and programs is not a trivial matter. For at least 20 years, foreign sales have accounted for approximately half of total sales for U.S. producers of films and programs.[4] In recent years the trade surplus on foreign sales of films and programs has probably exceeded a billion dollars.[1]

---

[1] See Braunstein and White (1985) for an analysis of antitrust problems raised by private, voluntary standard setting organizations.

[2] In this discussion we employ the FCC's definition of "advanced television" (ATV) as any television system which provides visual quality significantly greater than what is achievable with current commercial television systems in the United States. FCC (1988, fn. p. 4) The now commonly employed distinction between "high definition television" (HDTV) as television systems providing visual quality equivalent to 35 mm. film and "enhanced definition television" (EDTV) as television that is qualitatively improved relative to current television systems but somewhat short of the 35 mm. standard of reference is not relevant to the analysis presented below.

[3] Working Party 5 of the Planning Subcommittee of the FCC's Advisory Committee on Advanced Television Service estimate the present value of television receivers (excluding peripheral attachments such as VCRs) in U.S. homes at $22 billion. WP5 (1988, p. 28).

[4] See Renaud and Litman (1985).

The quality of films and programs viewed by American consumers also may be affected significantly by the foreign earnings of these entertainment products. The link between the variety and quality of video entertainment products and their potential earnings appears to be very strong. The possibility exists that various of the possible alternative technical standards for ATV may be incompatible in the sense that conversion of programs among standards for purposes of transmission may not be possible without significant loss of visual quality. The emergence of incompatible standards in different regions of the world would be a natural barrier to trade in films and programs. If this were to occur, the result could be less enjoyable home video entertainment for the American consumer.

This paper examines possible consequences for trade in films and programs of alternative choices among ATV standards. The analysis is based on a model of international trade in films and programs developed earlier by Wildman and Siwek (1987, 1988) to explain observed patterns in the international trade in these video products.[2] The model shows that if consumers have a natural preference to view films and programs produced in their native languages over foreign language productions, then the nature of commercial competition gives rise to an advantage in international trade for films and programs produced in the languages of the more populous and wealthy linguistic populations. The wealth and the size of the world's English-language population relative to other linguistic populations[3] thus explains the oft noted dominance in the international flows of programs and films of English-language productions, especially those from the United States and the United Kingdom.[1]

---

[1] U.S. Congress, Office of Technology Assessment (1986).

[2] The formal statement of this model is presented in Appendix B of Wildman and Siwek (1988). Less formal descriptions or extensions of the model are presented in Chapter 4 of the same book, and in Wildman and Siwek (1987). See also, Waterman (1988) for a similar model of trade in video entertainment programs.

[3] If the comparison is restricted to linguistic populations that reside primarily in countries that rely signficantly on market mechanisms for economic coordination, native English-speakers have nearly four times the purchasing power (measured by GNP) of the second (Japanese) and third (German) largest linguistic populations ranked by this measure. English speakers are more than three times as numerous as Japanese speakers and four times as numerous as German speakers. Hindi/Urdu and Spanish rank second and third to English in numbers of speakers, with about seven-eighths and five-eighths the English-speaking population respectively. However, both of these

The relevance of this model to the ATV standards debate is that differences in the ATV standards selected by various nations and groups of nations may have effects on trade in films and programs similar to the effects of linguistic differences if, as many believe will be the case, conversion of programs from one standard to another reduces their visual quality and entertainment value. (Henceforth, ATV standards for which conversion from one standard to another noticeably reduces visual quality will be referred to as incompatible standards.) Incompatible ATV standards would operate as barriers to trade, just as a loss in entertainment value when translating from one language to another is a barrier to trade. If current international alignments over ATV standards remain in place, the likely result of a layering of incompatible standards over existing linguistic divisions would be to reduce world trade in films and programs and to reduce the international sales of U.S. film and program producers.

ATV may also affect trading relationships by altering production costs in the film and television industries. ATV production techniques are already being employed for film and program production on a limited scale. Even if standard compatibility is not a problem, ATV may still affect trading relationships by altering the cost conditions that ultimately determine industry structures. The possible effects of changes in

---

linguistic groups rank relatively low in terms of purchasing power. Source: Wildman and Siwek (1988, Chapter 5).

[1] There are no simple and comprehensive measures of trade flows for films and programs that would permit straightforward comparisons among nations. Empirical descriptions of patterns of trade in films and television programs may be found in Wildman and Siwek (1988, chapter 2), Guback (1969, 1976), Varis (1984), and Varis and Nordenstreng (1974). These studies relied on a variety of measures that varied from country to country. However, the basic character of bilateral trading relationships in films and programs between the U.S. and other countries is illustrated by the fact that in 1986 U.S. exports of feature films to Japan were worth more than 35 times the value of Japanese film exports to the United States. Variety, December 9, 1987, p. 1. Japan, as a major producer of films, probably had a more favorable trade balance in films with the United States than most countries. The two Varis publications cited in this footnote calculate that the average nation imports programs that fill abount 30 percent of its television broadcast time, while for the United States the figure is between one and two percent.

production costs introduced by ATV on patterns of trade in films and programs will also be discussed from the perspective of the model of trade in video products. The basic economic model of trade in video products is described briefly in the next section and employed to examine the possible effects of ATV on patterns of trade in films and programs in Section III. Implications of the analysis for ATV trade strategy are discussed briefly in Section IV.

II. The economics of trade in films and programs

The model of trade in films and programs described in this section is a model of competition among commercially motivated film and program producers. Therefore, we begin by examining the nature of competition in the film and television industries. The economic distinction between "public goods" and "private goods" is critical to a proper understanding of competition in media industries. A pure public good is a product or service with the characteristic that consumption by one person does not reduce the amount available for others. That is, consumption of a public good is nonexclusive. National defense is a frequently cited example of a public good. All citizens of a country are protected to an equal degree, and, holding the defense budget constant, the amount of national defense consumed by each citizen is not affected by either increases or decreases in the national population. By contrast, a pure private good is a product or service for which consumption is entirely exclusive in the sense that consumption of a private good by one individual means that it cannot be consumed by others. Food products, for example, are private goods. The apple that I eat cannot be eaten by you.

Films and programs, like most products, have both public and private good components. The physical media, film and tape, on which performances are recorded are private goods. However, the performances recorded on these recording media are public goods in that viewing is nonexclusive and the size of the total audience for a given performance (the number of people consuming the same performance) may vary dramatically. Many economic analyses of competition ignore spending on the public good components of products, assuming implicitly that the cost and demand consequences of firms' expenditures on the public good components of products are small enough to be ignored without altering significantly the conclusions of the analysis.[1] For many products and services this assumption is warranted. For films and programs, however, the public good element cannot be ignored. The quality of the public good element is nearly the sole determinant of demand and sales, and the costs of the "creative inputs" (actors, directors, writers, special effects experts, camera crews, etc.) used to produce the public good that is the performance recorded on film or tape dominate the total of production and distribution costs.[2]

The decision of how much to spend on the creative inputs that make up a production budget is critical to the financial success of a film or program. Assuming that there is a positive relationship between expenditures on creative inputs and audience appeal, spending more on creative inputs will increase the revenues *likely* to be generated by a program or film.[3] If anticipated revenues increase by more than a dollar for a dollar increase in the production budget, expected profits are increased by doing so. The profit maximizing producer should continue to increase his production budget as long as the anticipated increase in revenue exceeds the increase in budget required to produce it. Of course there is a limit to how much the expected profits of a film or program can be increased by increasing the production budget. At some point, the increase in revenue generated by a given small increase in the budget will begin to decline. Eventually, the point will be reached where an extra dollar spent on creative inputs will generate less than a dollar increase in revenue. Expected profits will be maximized at that budget for which an extra dollar spent on creative inputs increases expected revenue by exactly a dollar. To increase the budget beyond this point would reduce profits.[1] These relationships are illustrated in Figure 1.

---

[1] This is not true of analyses that incorporate the effects of advertising or R&D, since advertising and R&D have important public good characteristics.

[2] According to the Motion Picture Association of America (MPAA), in 1987 the negative costs for features produced by MPAA member studios averaged $20 million while the costs of prints and advertising for the same films averaged $9 million. Variety March 2, 1988, p. 3.

[3] Among other things, "spending more on creative inputs" would include increases in budgets to hire actors and actresses with more demonstrated ability to draw audiences, hiring directors and writers with better reputations, and spending more on special effects. Increases in these types of expenditures would increase the likely audience appeal, and thus, the audience and revenue generated by a film or program. However, given the uncertainties inherent in creating films and programs to match audience tastes, the relationship between expenditure on creative inputs and audience appeal is only a probabalistic one. There is no guarantee that the use of higher priced talent will increase the audience for a particular film or program.

The curve labelled R in Figure 1 illustrates the relationship between the size of a production budget and expected revenue for a film or program. The intersection of R with the horizontal axis indicates the minimum budget required to attract any viewers. A film or program with a budget smaller than this will have no audience. R rises rapidly at first because at smaller budgets increases in spending on creative inputs have a large impact on expected earnings. R gradually becomes flatter with increases in the production budget, indicating that for films and programs, as with other products, diminishing returns limit the extent to which production expenditures profitably can be increased.

Production costs are translated from the horizontal to the vertical dimension by the 45 degree line in Figure 1 to facilitate the comparison of revenue and cost. Profits for the imaginary firm depicted are represented by the difference between R and the 45 degree line. Expected profits (the distance between R and the 45 degree line) are maximized with the budget at which R is parallel to (has the same slope as) the 45 degree line. This budget is indicated by B* in Figure 1. At B* a one dollar increase in the production budget would generate exactly a one dollar increase in revenue. Beyond this point further expenditures on creative inputs will continue to increase revenue, but by less than the cost of doing so.

Implicit in the revenue-budget relationship (R) shown in Figure 1 are assumptions about the size of the potential audience (say the number of tv households or the number of potential theater goers) and the number of other producers competing for that audience. Holding the number of competitors constant and increasing the size of the potential audience should raise R as illustrated by the movement from R_0 to R_1 in Figure 2. Holding the size of the audience constant and increasing the number of competitors should lower R, as indicated by the movement from R_0 to R_2 in Figure 2. Note that the profit maximizing budget increases if R rises and declines if R falls.

---

[1] Note that this description of the determination of the profit maximizing budget assumes risk neutrality on the part of the producer. Risk aversion would lead to a budget smaller than the budget at which expected profits are maximized, since the producer would accept a trade-off of lower expected profits for less risk.

Because expected profits are positive, none of the situations depicted in Figure 2 could represent the long run expectations of firms in a competitive market. Positive profits would attract new competitors until expected profits were zero, in which case revenue functions would be just tangent to the 45 degree line. Two such tangencies are shown in Figure 3. R_L is a revenue-budget relation for a large market and R_S is a revenue -budget relation for a smaller market. Note that the tangency of the revenue-budget line with the 45 degree line for the large market is to the right of the tangency for the small market, indicating larger production budgets in large markets than in small markets.[1] The competition for shares of audience and revenue leads producers to spend more on creative inputs in larger markets than in smaller markets. This is a direct concequence of the public good characteristics of recorded video entertainment.

To see why production budgets should be larger in large markets than in small markets, assume that for a single market the number of producers is held constant while the number of viewers in the market increases. Production budgets that previously had been in equilibrium will now have to be adjusted. Whereas before the increase in viewers, the expense of hiring slightly higher quality creative inputs had just equaled the increase in revenue likely to be generated, now the number of viewers affected and the revenue they represent is larger. This means that a small increase in the production budget is likely to increase revenue by an even greater amount. Producers would thus begin to increase their production budgets. Of course, the fact that all producers can be expected to increase their production budgets will reduce to some degree the gain to any one of them from doing so, but not by enough to totally offset the effects of increased market size. After all adjustments have been made, production budgets should be higher in the new equilibrium.[2]

The number of producers is also likely to increase in response to an increase in market size. For the individual producer, this will also tend to reduce the returns to increased spending on creative inputs. However, if, as seems reasonable, viewers value both the diversity provided by having a larger selection of films or programs and the improvements in the entertainment value of a given program or film that can be produced by spending more on creative inputs, larger markets should provide more of both than smaller markets.

---

[1] See Wildman and Siwek (1988, Appendix B) for the formal derivation of this result.

[2] Higher quality competition is the reason why the intercept of R_L with the horizontal axis is to the right of the intercept of R_S.

Opening of trade between the two markets illustrated in Figure 3 will affect the equilibria described in several ways. First, since producers in both markets are competing for audience and revenue in a larger combined market, equilibrium budgets should rise. If the distinction between the two markets is entirely political, so that on economic grounds there would be no reason to distinguish between the two, then trade should lead to a complete equalization of budgets. On the other hand, if differences in language, culture, or tastes (or ATV standards) produce a natural differention between the video products of the two countries such that viewers have a natural preference for home country films and programs, production budgets in the larger country will continue to be somewhat larger than budgets in the smaller country.

This model of competition and trade in films and programs suggests that if, as one would intuit, viewers strongly prefer films and programs in their own languages, we should observe countries whose populations belong to large and wealthy natural language populations producing larger budget films and programs than countries whose citizens speak the languages of smaller linguistic populations. These types of relationships are documented in Wildman and Siwek (1988) and Waterman (1988).[1] It is the greater expenditures on creative inputs by producers in larger markets that give rise to what has been described as a "one-way flow" of programs and films from large and wealthy countries (mainly English-language countries) to smaller, poorer countries. The citizens of larger and wealthier countries are generally members of the larger and wealtier linguistic populations. The greater entertainment value purchased with larger budgets in the larger, wealthier countries offsets to some degree the native language preference for home country productions in smaller countries with different languages. Films and programs from smaller countries seeking audiences in larger countries have the handicap of smaller budgets in addition to native languages preferences to contend with.

---

[1] Figures reported by Wildman and Siwek (1988) show that for 1982-83, average budgets for U.S. feature films were four to five times average budgets for the major film producing nations of Japan, France and Italy. Waterman (1988) reports similar relationships for television budgets.

III. Possible Effects of ATV on Trade in Films and Programs

ATV has the potential to affect trade in films and programs in at least two ways. (1) If the standards selected by various countries or groups of countries are incompatible, so that programs produced to take maximimum advantage of the capabilities of one standard suffer from quality degradation when converted for use with a different transmission standard, the effect may be similar to the effects of linguistic differences among consumer populations. Compatibility problems could thus alter the flows of trade in films and programs. (2) High definition production techniques may alter the relationship between production expenditures and the quality of films and programs as perceived by viewers. Depending on the nature of such cost effects, the advantage in trade currently realized by producers producing in the languages of large linguistic populations could be either enhanced or diminished.

The Compatibility of Prospective ATV Standards

Television standards are technical specifications that govern both the production of television programs and the transmission of programming to viewers' sets. The viewing quality of the programming transmitted is maximized when the technical characteristics of the production standard are well matched to the technical characteristics of the transmission standard. The transcoding of a program for transmission may result in a significant loss of picture quality if production and transmission standards are not well matched. Some quality loss in transcoding is inevitable. Whether quality degradation is substantial or not depends on the standards involved. Currently the world is divided into three blocks of countries using the NTSC, PAL, and SECAM standards. Problems of quality degradation associated with converting programs among these standards have not been so great as to limit seriously international trade in programs, although occasionally the cost of conversion does prevent trades in marginal programs.

At this time, it appears likely that when ATV has finally supplanted the current conventiol television systems, the world will once again be divided into blocks of countries employing different production and transmission standards.[1] Japan has committed to a production standard with 1125 lines per frame and 60 fields per second (1125/60), has developed a DBS transmission system based on standards well suited to 1125/60 production, and is working on terrestrial transmission systems compatible with its 1125/60 production standard.[2] While there is considerable support for establishing 1125/60 as a worldwide production standard to facilitate international program exchanges, the chances for adoption of a single worldwide production standard are diminishing. Europe appears to be moving toward adoption of the Eureka transmission system which incorporates a 1250/50 standard. To this point, Europeans have rejected attempts to establish 1125/60 as a worldwide production standard, claiming that conversion of 1125/60 programs for Eureka transmission would be too difficult. In the U.S., the FCC has initiated an ATV inquiry, but it is still too early to identify favorites among many potential candidates for production and transmission standards. There is strong support for a 1125/60 production standard, but 1125/60 also has strong detractors who argue, among other things, that a 1125/60 production standard would not be especially compatible with the NTSC transmission system now used in the U.S., which operates at 59.94 frames per second. NBC recently introduced an alternative 1050/59.94 production standard designed with NTSC compatibility in mind. The NBC system already has won some powerful endorsements.[3] The FCC's tentative decision to require NTSC compatibility would appear to favor the NBC standard.[4]

At this point we don't know just how much signal imparement will be produced in transcoding 1125/60 programs for Eureka transmission or for transmission with a NTSC compatible ATV system. Proponents of a 1125/60 world standard say that any imparement will be neglible. By contrast, supporters of the new NBC production standard claim that transcoding a 1125/60 production for transmission on a 59.94 system would "produce unacceptable picture artifacts" in high definition pictures, although there would be no serious problem for standard NTSC resolution.[1] There is a similar division of opinion concerning the signal degradation likely to be produced in converting 1125/60 productions for Eureka transmission.

---

[1] The analysis that follows assumes a world in which all television is some form of ATV. Developments during a period of transition from current conventional systems to ATV systems are not considered.

[2] The various members of the MUSE family of transmission standards.

[3] Besides NBC, other companies endorsing the 1050/59.94 standard include Capital Cities/ABC, North American Phillips Corp., Zenith, Tribune Broadcasting Co., Faroudja Laboratories Inc., Thomson Consumer Electronics, and the David Sarnoff Research Institute. Broadcasting October 17, 1988, p. 31.

[4] FCC (1988).

Trade with Incompatible Standards

Assuming that different countries and regions do select different transmission systems, as now appears all but certain, and assuming further that producers in these countries and regions choose production standards that maximize the benefits of domestic transmission systems, what are the likely consequences for international trade in films and programs of serious standards incompatibility? That from an analytical perspective, the effects on trade of incompatible ATV standards should be much the same as the effects of linguistic differences has already been mentioned. A program translated from one language to another, whether in the viewer's mind, or in a studio via subtitling or dubbing, has less entertainment value in the language it is translated into than in the language in which it was produced. Analogously, if conversion from one ATV standard to another involves a noticable loss of signal quality, the entertainment value of programming will be diminished by conversion. Thus, just as there are now communities of viewers distinguished by preferences for programs in their native languages, in the future these linguistic divisions among audiences may be overlaid by additional divisions based on ATV production and transmission standards. If standards based coversion problems are significant, the result will almost certainly be a reduction in world trade in television programs overall since the attractions of imported programs will decline. However, for a block of countries, such as Europe, separated from the United States by a common set of ATV standards, the effect may be increased trade in programs and films within the block, if U.S. productions are made less attractive by incompatible standards.[2]

---

[1] Comment of Julius Barnathan, ABC president of broadcast operations and engineering, as reported by Broadcasting October 17, 1988, p. 31.

[2] If films continue to be produced primarily in 35 mm., then the effects on films would be much smaller. If, on the other hand, films were produced in high definition and distributed to theaters via DBS, as has been done in Japan, the consequences for trade in films would be similar to that in programs.

U.S. producers would almost certainly be hurt by incompatible international standards. U.S. films and programs are such a large fraction of films and programs traded, that a decline in the total would almost certainly mean a decline for U.S. producers. Probably of greater significance would be the fact that the U.S. would be separated by differences in standards from its two largest foreign markets outside of North America, Europe and Japan. In 1984 Europe and Japan accounted for over $330 million of the $654 million in foreign theatrical revenues generated by the U.S. major studios.[1] Comparable figures for foreign television sales are not available, but U.S. sales of programs in foreign markets are now variously estimated at between $500 million and $1 billion.[2] It is likely that Japan and Europe account for a similar fraction of U.S. foreign program sales. Italy alone paid nearly $89 million for U.S. programs and films to be shown on television in 1983.[3]

A second adverse affect from the U.S. perspective would be a splintering of the world English language population according to differences in high definition standards. Of the 409 million native English speakers worldwide, 234 million live in the United States and may be considered somewhat "captive" to a U.S. standard. Better than 59 million, over one seventh of the total, live in the United Kingdom and Ireland, and presumably would adopt Eureka standards.[4] The large population of Europeans who speak English as a second language would also be "captive" to Eureka standards. Therefore, the "comparative advantage" currently enjoyed by American film and television producers by virtue of the size and wealth of the world's English speaking population could be reduced considerably.[5]

According to the trade theory described in Section II, production budgets increase with the size of the potential audiences for films or programs. A dimished world market for U.S. productions would mean lower budget films and programs for American viewers.[1]

---

[1] Variety June 12, 1985, p. 3.

[2] See, for example, "American TV Tightens Its Grip on the World," Channels of Communication January/February 1984, pp. 27-30.

[3] Variety May 9, 1984, p. 303.

[4] Sources: The World Almanac and Book of Facts (New York: Newspaper Enterprise Association, Inc., 1984); World Bank, World Tables, 3d ed., vol 1 (Baltimore, Md.: Johns Hopkins University Press, 1983).

[5] Producers in the U.K. might be hurt the most, since their work would no longer be as attractive to the very large American market. British tv producers had foreign sales of $188 million in 1984-85, 59 percent of which was to the United States. Variety August 10, 1988, p. 1.

Trade Consequences of Changes in Production Costs Following HDTV

Independent of the effects of standards compatibility on trade, ATV may also affect trade through its effects on the costs of producing programs and films. However, there is no definitive answer to the question of what the effects will be on trading relationships. Producers will now have a cheaper alternative to 35 mm. film for acheiving visual quality at or close to the 35 mm. standard. That more visual quality can be acheived for less would appear to favor producers in smaller markets that produce small budget films and programs. Many films in Europe currently are produced in 16 mm., which is both less expensive and of lower quality than 35 mm. Some of these films might now be shot with ATV techniques. Again, this would favor small market producers. On the other hand, we do not know how ATV technology will affect the demand for complementary creative inputs such as acting talent. If other creative inputs become more valuable, total budgets may actually increase, which would favor large market producers. At this point there is no way to predict with certainty just who will be favored by the cost effects of high definition technology.

IV. Implications

The clearest implication of the preceding analysis is that incompatible world ATV standards hold the potential for considerable harm to the U.S. industry that produces television programs and films. Given that standards adopted by other countries are largely beyond the influence of U.S. policy makers, some consideration should be given to selecting ATV standards that minimize incompatibility problems with the standards adopted by major trading partners.[2] This raises the possibility of actually enhancing the competitive position of U.S. program producers. If U.S. ATV standards were relatively compatible with the standards of important trading partners whose standards were not compatible with each other, only U.S. productions would be unhindered by standards incompatibility. There would then be an incentive for foreign producers trying to reach international markets to adopt U.S. production standards, which in turn might benefit U.S. hardware manufacturers. Thus the trading interests of both hardware and software producers might be served. The risk of such a "wait and see" strategy is that foreign standards might become established as *de facto* standards in the U.S. for lack of a domestic alternative.

---

[1] Again, note that to the extent that motion pictures remain on a 35 mm. standard, the effects of ATV standards on the market for films would probably be considerably less, even though television is now a major purchaser of theatrical releases.

[2] Assuming such a standard is feasible.

References

Braunstein, Y. M., and White, L. J., "Setting Technical Compatibility Standards: an Economic Analysis," The Antitrust Bulletin, Summer 1985, p. 337.

Caranicas, P., "American TV Tightens Its Grip on the World," Channels of Communication, January/February 1984, p. 27.

Cohn, L., "It's Still Paying to Play the Movies: Majors Recouped Costs on '87 Pix," Variety, March 2, 1988, p. 3.

FCC, "Tentative Decision and Further Notice of Inquiry," MM Docket No. 87-268, September 1, 1988.

Guback, T. R., The International Film Industry: Western Europe and America Since 1945, (Bloomington: Indiana University Press, 1969)

______, "Hollywood's International Market," in The American Film Industry, ed. Tino Balio (Madison: University of Wisconsin Press, 1976), p. 387.

Murphy, A. D., "Top Markets for U.S. Pics, 1984" Variety, June 12, 1985, p. 3.

"NBC Unveils New HDTV Standard," Broadcasting, Oct. 17, 1988, p. 31.

Segers, F., "Best Customer for Pics From Japan Is U.S.," Variety, December 9, 1987, p. 1.

Renaud, J. L., and Litman, B. R., "Changing Dynamics of the Overseas Market Place for TV Programming," Telecommunications Policy, September 1985, p. 249.

U.S. Congress, Office of Technology Assessment, "Trade in Services: Exports and Foreign Revenues Special Report," OTA-ITE-316 (Washington D.C.: U.S. Government Printing Office, September 1986), p. 89.

Variety, May 9, 1984, p. 303.

Varis, T., "The International Flow of Television Programs," Journal of Communications, Winter 1984, p. 143.

_____, and Nordenstreng, "Television Traffic, A One-Way Street," (Paris: Boudin/UNESCO, 1974).

Waterman, D., "World Television Trade: The Economic Effects of Privatization and New Technologies," Telecommunications Policy, June 1988, p. 141.

Wildman, S. S., and Siwek, S. E., "The Privatization of European Television: Effects on International Markets for Programs," The Columbia Journal of World Business, Fall 1987, p. 71.

______ and_____, International Trade in Films and Television Programs, (Cambridge: Ballinger, 1988).

Working Party 5, "Economic Factors and Market Penetration", report to the Planning Subcommittee of the FCC Advisory Committee on Advanced Television Service, May 9, 1988.

FIGURE 1

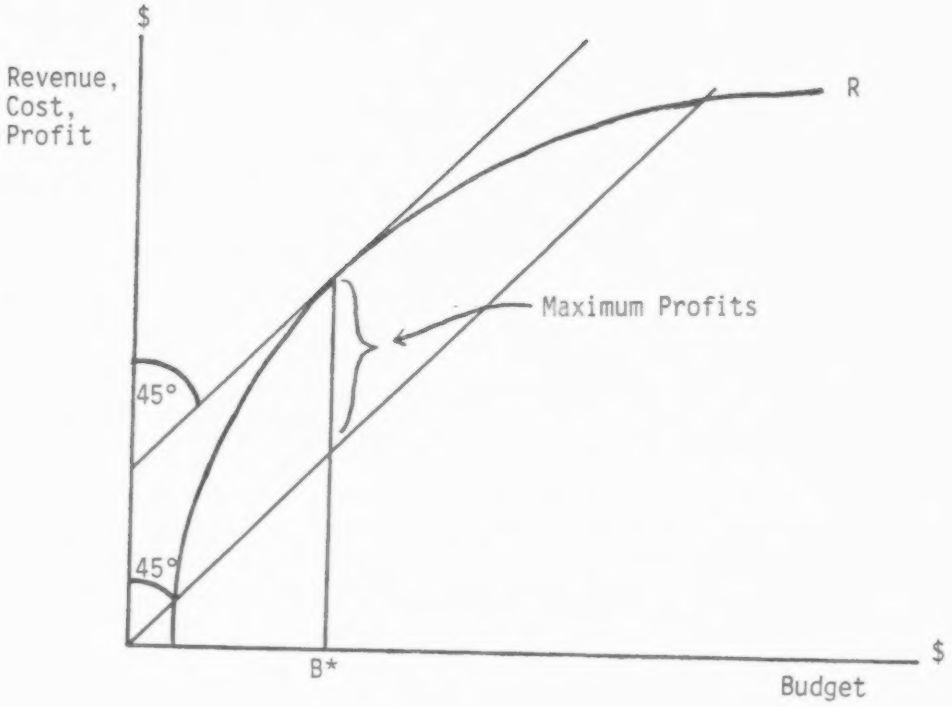

$
Revenue,
Cost,
Profit
R
Maximum Profits
45°
45°
B*
$
Budget

FIGURE 2

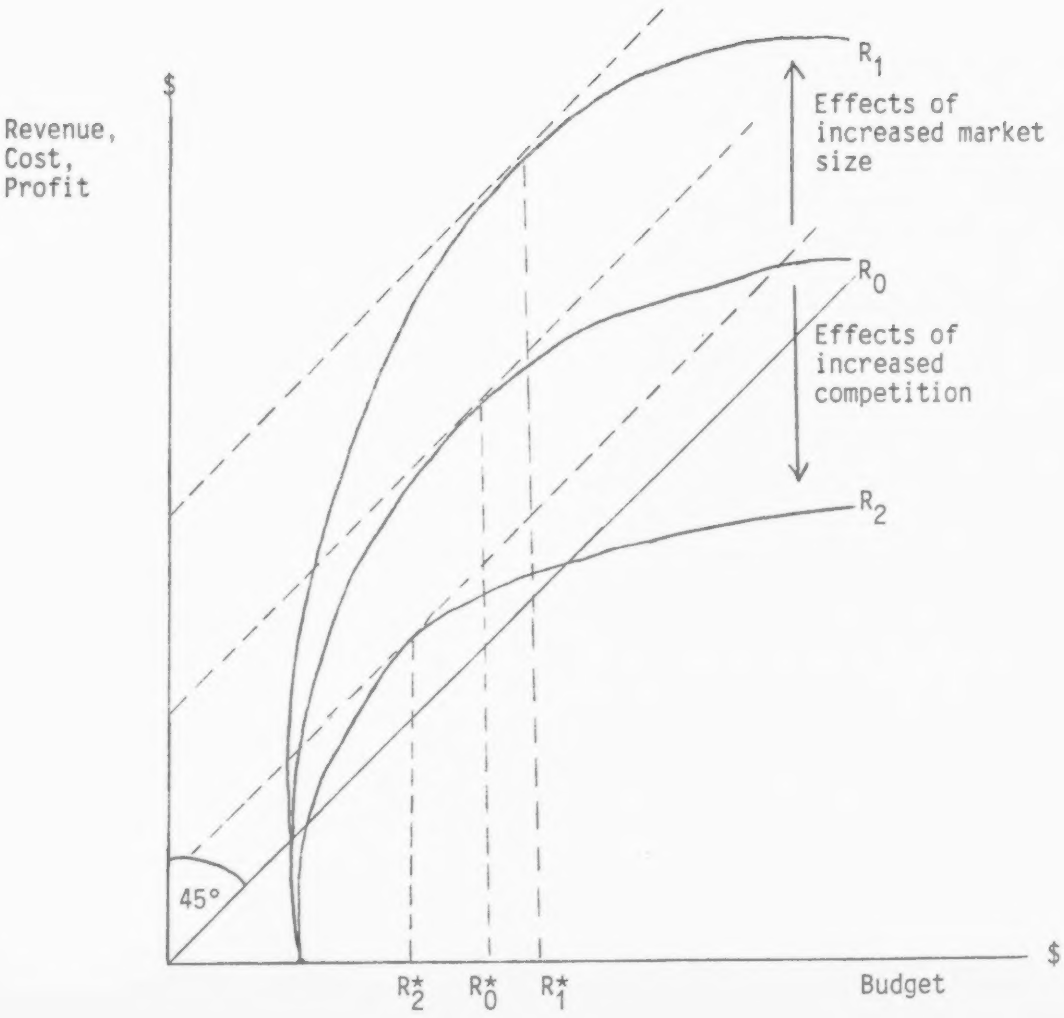
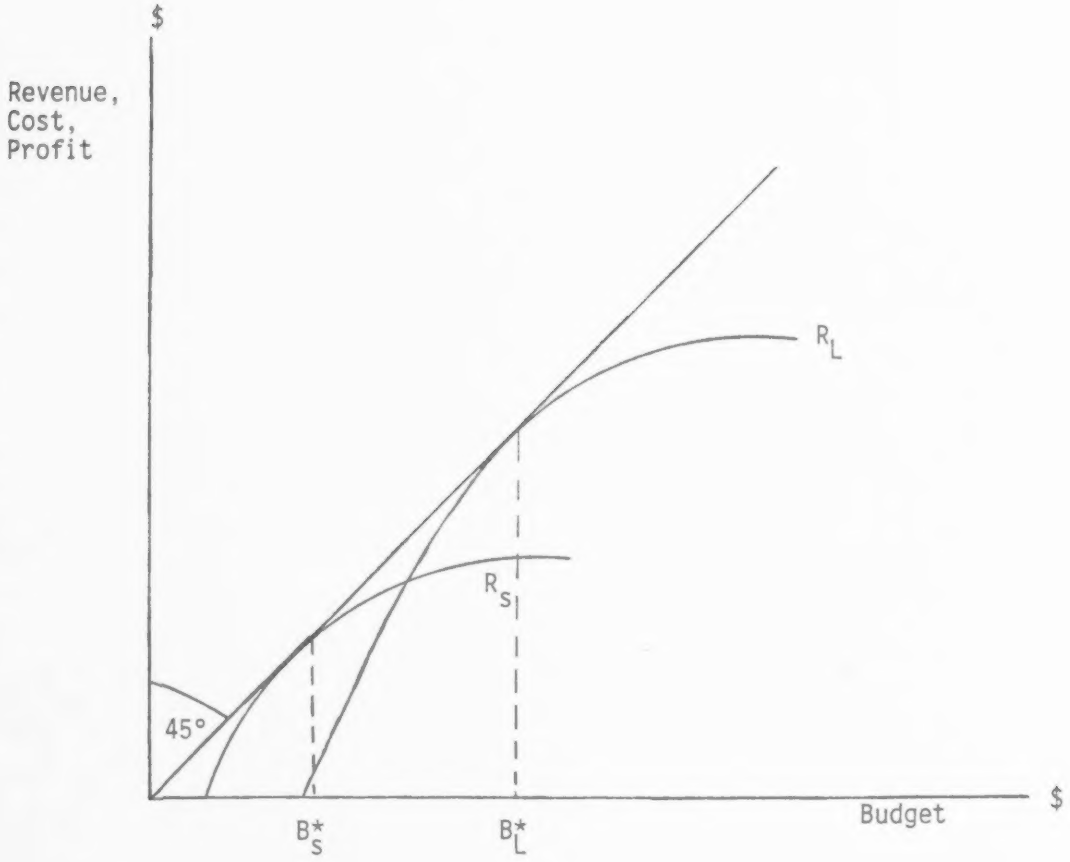

$
Revenue,
Cost,
Profit
R1
Effects of increased market size
R0
Effects of increased competition
R2
45°
$
Budget
R*2
R*0
R*1

FIGURE 3



Carnegie Mellon University
Department of Engineering and Public Policy
Pittsburgh, Pennsylvania 15213

Estimation of A Market Growth Model with Network Externalities and Dynamic Pricing Using an Extended Kalman Filter

by
Jinhong Xie
and
Marvin A. Sirbu

Presented at the
1988 Telecommunications Policy Research Conference

Airlie, Virginia
October 31, 1988

This research was supported in part by grants from Bell of Pennsylvania and from the Nynex Corporation.

1. Introduction

Network externalities are an essential feature of most telecommunication markets and have an important bearing on pricing decisions, especially in the case of a new product or service. At product introduction, network externality effects are weak, and the product may fail to reach "critical mass." An optimal penetration pricing strategy may require the vendor to price below cost initially, raising the price as the growth of subscribers raises the value of the network externality. If the price is correctly varied over time, a monopoly supplier can attract a stream of new adopters and thereby ensure the growth of the market and maximize the present worth of future profits. If the demand curve and diffusion characteristics of the adoption process are known, Dhebar and Oren (1985) have shown that the penetration pricing problem can be formulated as an optimal control problem in which the objective is to determine not a single optimal price, but a complete price trajectory over time. They show that, in general, the optimal strategy is one which starts at a price of zero when there are zero subscribers, and increases the price monotonically until the profit maximizing equilibrium is reached. Losses during the early phases of product introduction are to be recovered after the market reaches critical mass.

Unfortunately, it is not possible to accurately establish, in advance, the demand and diffusion parameters which determine the optimal price trajectory. Following an incorrect pricing path based on a faulty *a priori* estimate of the demand may lead in the best case, to lost profit opportunities from pricing too low, and, in the worst case to a total collapse of the enterprise if overpricing results in failure to reach critical mass. Incorrect estimates of the rate new adopters will come on stream can lead in the best case to a shortage of capacity and lost revenue, and in the worst case to over-investment and exhaustion of capital before the market becomes self-sustaining.

In principle, a careful analysis of the market response to the early stages of the price trajectory can be used to verify a priori assumptions regarding the demand and diffusion parameters. The problem of predicting future new product sales from early data has been well studied. It is difficult enough in the usual case—without network externalities, and with price held constant—to estimate the parameters of the diffusion process from early sales data.

The present paper shows how an Extended Kalman Filter can be used to continuously update estimates of the underlying demand and diffusion parameters of the market. These updated estimates are then used to continually revise the pricing strategy resulting in significant improvements in profitability. Our formulation of the estimation problem is quite general, and the estimation technique could as well be applied to other markets characterized by a non-linear diffusion equation with a set of parameters to be estimated.

2. Diffusion and Pricing in a Market with Network Externalities

Rohlfs (1974) was the first to characterize the nature of demand in markets with a network externality. He observed that such markets may exhibit what appears to be an upward sloping demand curve: as more subscribers join the network, the value of the network is increased, and thus the willingness to pay. (Eventually, of course, the demand curve must turn down.) Dhebar and Oren (1985, 1986) were the first to characterize the evolution of subscription demand in a market for a new product with network externalities. By combining the demand function of Rohlfs as a description of potential demand, with a diffusion equation to describe the process by which subscribers actually sign up, they created a dynamic model of network growth. We summarize their model here, before discussing its limitations.

Following Rohlfs, Dhebar and Oren assume an infinitely large population of heterogeneous consumers, each of whom has a constant positive income. Consumer preferences are stationary and are represented by an instantaneous willingness-to-pay function $W(\eta, x)$, which is individual η's reservation price for subscription, given a subscriber set $[0,x]$. It is assumed that :

1. η is uniformly distributed over the interval $[0, 1]$.
2. $W(\eta, x)$ is continuously differentiable and monotonically decreasing in η and increasing in x for any $x \in [0, 1]$.
3. Nonsubscribers always join the network in sequence of increasing η.[1]

Given a subscription price p, and assuming that there are no costs associated with either the initiation or the renunciation of subscription, a consumer η will subscribe if and only if

$$W(\eta, x) \geq p$$

Thus, the three determinants of a consumer's subscription decision are his index, the subscriber set size and the subscription price. Of these, the consumer index is assumed to be time-invariant. However, consumer expectations about future subscriber set size may be considered in the decision to subscribe.

Assume, for example, that the market response is not instantaneous. Then the actual number of subscribers at any point in time may be less than the "potential demand". At one extreme, subscription decisions may be based on the size of the existing subscriber set $[0, x]$ as if its size will remain unchanged. But, if the subscription demand differs from the existing subscriber set, then the size will change and consumer η may decide not on the basis of $[0, x]$, but on the basis of a network extending over (at least) $[0, \eta]$. Real consumer behavior may not correspond to either of these extremes, i.e., complete myopia about subscriber set growth or perfect growth anticipation. Network growth will only be partially anticipated. To take into account the possibility of imperfect knowledge about the exact nature of consumer ordering, Dhebar and Oren introduce a growth anticipation parameter $\alpha, 0 \leq \alpha \leq 1$. It is assumed that each consumer η bases his decision on an anticipated subscriber set $[0, \alpha\eta + (1-\alpha)x]$. Then, the total subscription demand is $d^{\alpha}(x, p)$, where

$$d^{\alpha}(x,p) = \max\{\eta:\ 0 \leq \eta \leq 1,\ W(\eta, [\alpha\eta + (1-\alpha)x]) = p\} \qquad (2\text{-}1)$$

and the set of consumers wishing to subscribe is $[0, d^{\alpha}]$.

Network growth occurs only when the subscription demand exceeds the existing network size: $d^{\alpha}(x, p) > x$. The expansion will stop when $d^{\alpha}(x,p) = x$. There are different diffusion models to describe the growth of the market . The general diffusion model could be viewed as:

$$\frac{\partial x}{\partial t} = G(d, x(t)) \qquad (2\text{-}2)$$

where $x(t)$ is the cumulative number of "adopters" (the current market size), and d is the potential market for the new product and G is an arbitrary function. In the present case, $d = d^{\alpha}(x, p)$ and the general diffusion model (2-1) yields a revised diffusion model in which price can be used as a control variable.

---

[1] It is this assumption which allows the set of subscribers to be characterized as the set from $[0, x]$.

$$\frac{\partial x}{\partial t} = G\,(d^{\alpha}\,(x, p), x) \qquad (2\text{-}3)$$

Dhebar and Oren go on to show how one may solve for an optimal pricing strategy over time, $p^*(t)$ designed to maximize the present value of future profits, assuming a discount rate δ, and a cost function $C(x)$. They show that the optimum trajectory is dependent only on x. Thus,

$$p^*(t) = p^*(x(t)) \qquad (2\text{-}4)$$

They also provide a detailed illustration of how to calculate $p^*(x)$ for the case where

$$W\,(\eta, x\,) = W_O\, x\,(1 - \eta) \qquad (2\text{-}5)$$

$$G\,(d, x) = \ b\,(d^{\alpha}\,(x, p)\text{-}x) \qquad (2\text{-}6)$$

and

$$C(x) = C_O x \qquad (2\text{-}7)$$

We will return to the calculation of $p^*(x)$ in section 5.

2.1 Practical difficulties in implementing optimal dynamic pricing.

The construction of the optimal price trajectory assumes that the vendor has perfect knowledge of both $W(\eta, x\,)$ and $G(d, x)$, *e.g.* a priori knowledge of the parameters W_O and b. In practice, of course, the vendor has very limited information a priori on which to base estimates of these parameters.

If the estimates used to calculate the pricing rule $p^*(x)$ are not correct, there can be a number of unfortunate outcomes. Overestimating W_O leads to a pricing rule $p^*(x)$ which is too high, with the result that the subscriber base never reaches critical mass, and the enterprise collapses. Too high an estimate for the diffusion parameter b leads to overoptimism about the time required to reach equilibrium, with the result that the vendor may build up capacity too quickly, or run out of capital before becoming profitable. In a regulated environment, failure to meet profitability projections may lead the regulators to conclude that a strategy of penetration pricing to be repaid out of future profits was nothing more than a ruse to disguise a cross-subsidy to the new service from other ratepayers.

An underestimate of W_O will lead the vendor to price too low. The result will be an equilibrium at a larger overall network size, but at a lower price with correspondingly lower profits.

3. Estimation of Model Parameters.

Clearly, then the vendor would like to use early sales data to refine his a priori estimates of the critical parameters, and modify his pricing policy accordingly.

The diffusion model given by (2-3) has three important characteristics:

1. It is highly non-linear because of the complexity of $d^{\alpha}(x,p)$ and G.
2. While the model is described by a continuous differential equation, observations of sales data used to update the parameter estimates will occur only at discrete intervals.
3. Although we have assumed stationarity, the coefficients in the model may well be time-varying.

3.1 Comparison of Estimation Methods

In recent years, a number of estimation procedures have been suggested for estimating the parameters of diffusion models. Ordinary Least Squares (OLS) regression and Maximum Likelihood Estimation (MLE) are the most widely used approaches. Others, concerned that the model parameters may in fact be time varying, have proposed the use of adaptive estimation procedures including adaptive filter (AF) approaches and the discrete Kalman Filter (KF).

All of these approaches have been applied to the problem of estimating variants of the well-known Bass diffusion model for new product sales. If $N(t)$ is the cumulative number of adopters at time t, then the Bass model leads to the following diffusion equation

$$\frac{\partial N(t)}{\partial t} = (p + \frac{q}{m} N(t))\ (m - N(t)) \tag{3-1}$$

where p and q are unknown diffusion parameters referred to as the coefficient of innovation and the coefficient of imitation, respectively, and m is the maximum number of potential adopters, and is assumed constant. (This corresponds to $d=m$ in equation (2-2)).

OLS was suggested by Bass in his original exposition of the model (Bass, 1969). He formulated a discrete analog of equation (3-1) in terms of $X(i)$ the number of new adopters in the ith period:

$$\begin{aligned} X(i) &= (p + \frac{q}{m} N(i\text{-}1))\ (m - N(i\text{-}1)) + \varepsilon(i) \\ &= pm + (p - q)\, N(i - 1) - \frac{q}{m} N^2(i - 1) + \varepsilon(i) \\ &= \alpha_1 + \alpha_2 N(i - 1) + \alpha_3\ N^2(i - 1) + \varepsilon(i) \end{aligned} \tag{3-2}$$

where $\alpha_1 = pm$, $\alpha_2 = (p - q)$, $\alpha_3 = -q/m$, and $\varepsilon(i)$ is the error. Estimates of the coefficients α_1, α_2, α_3 can be obtained by OLS regression, and the parameters p, q and m derived therefrom. The transformation of the variables p, q, and m into α_1, α_2, and α_3 is necessary to produce a linear equation suitable for simple OLS estimation. The MLE method described by Schmmittlein and Mahajan (1982), is applicable only when the cumulative distribution function of the number of adopters can be expressed as an explicit function of time, t. Moreover, even when such an explicit function is available, finding the parameter values which maximizes the likelihood function requires the use of heuristic search methods and is computationally inefficient. Finally, the method they describe can only be used to estimate constant parameters.

Several authors have suggested the use of Adaptive Filters (Makridakis and Wheelwright 1977, 1978, Ledolter and Kahl 1984) or Kalman Filters (Kahl and Ledolter 1983, Meade 1985) for estimating the parameters of market growth processes. Typically, the authors posit autoregressive models to describe the dynamics of growth:

$$Z_t = \sum_{i=1}^{k} \beta_i Z_{t-i} + \varepsilon_t \qquad (3\text{-}3)$$

where $\mathbf{Z}_t = \{Z_{t-k}, \ldots Z_t\}$ is the last k sequential observations of the cumulative number of adopters, and $\beta = \{\beta_i\}$ i =1, 2, ... k are the unknown parameters. Both of these methods are applicable to the case of time-varying parameters, β_t. The difference between the two methods is in the way they update estimates of the unknown parameters. For the adaptive filter approach, $\hat{\beta}_t$, the estimate of the coefficient vector $(\beta_1, \beta_2, \ldots \beta_K)'$ at time t is updated according to

$$\hat{\beta}_t = \hat{\beta}_{t-1} + 2\,\alpha\,\mathbf{Z}_t\,(Z_t - \mathbf{Z}_t'\,\hat{\beta}_{t-1}) \quad (3\text{-}4)$$

The constant α is a learning constant and determines how much weight should be given to the most recent one-step-ahead forecast error

$$Z_t - \mathbf{Z}_t'\,\hat{\beta}_{t-1}$$

In the Adaptive Filter procedure, this constant must be supplied by the analyst. However, for the Kalman Filter, the weight or gain $K(t)$ is calculated from the observations.

All of the above papers start by assuming a discrete form of the diffusion model, and use an autoregressive formulation. However, the complexity of demand, $d^{\alpha}(x, p)$, in the problem at hand makes it very difficult to express the diffusion process in an autoregressive form. Furthermore, a time interval bias is introduced by discretizing the continuous differential equation which describes the diffusion process. Since the purpose of estimating the diffusion model here is to be able to update a continuous optimal pricing function, this bias could be significant.

As discussed in section 2.1, we require a method capable of estimating—potentially time-varying—parameters of a complex, non-linear, continuous differential equation. Recent developments in optimal estimation theory have led to the development of a method which is well-suited to these characteristics—an Extended Kalman Filter with continuous state equation and discrete observation equation (EKF(C-D); Lewis, 1986). The resulting algorithm is easy to use and can be shown to provide the best linear estimate in the presence of arbitrary disturbances, and an optimal estimate if the noise is Normally distributed. We do not need to change the diffusion model from a differential equation to a difference equation, nor do we need a complete description of the observation statistics as required by the MLE; we need know only the mean and covariance of the noise processes. As long as the diffusion process can be expressed by a differential equation with the unknown parameters, these parameters can be estimated from discrete observations using a simple sequential algorithm. The algorithm makes effective use of any prior distribution on the parameters; and, finally, it produces not only an estimate for the parameter but a covariance matrix for the estimates.

4. The Extended Kalman Filter for Continuous State and Discrete Observations

The Extended Kalman filter (EKF) is an approximate filter for nonlinear systems. The EKF (C-D) method described here is a method for estimating, from discrete observations, the parameters as well as the state for a system described by non-linear continuous differential state equations.

EFF (C-D) assumes that the nonlinear model can be expressed in the form:

$$\frac{\partial \boldsymbol{x}}{\partial t} = G\,(\,\boldsymbol{x}(t), \boldsymbol{u}(t), \theta, t) + \boldsymbol{M}\boldsymbol{w}(t) \tag{4-1}$$

where $\boldsymbol{x}$ is the state vector, $\boldsymbol{u}$ is the control vector, θ is the unknown parameter vector, G is a nonlinear function of $\boldsymbol{x}$, $\boldsymbol{u}$, θ and t, $\boldsymbol{w}$ is the process noise vector and $\boldsymbol{M}$ is a known matrix.

When the EKF method is used for parameter estimation, it is necessary first to include the unknown parameters in the state vector. Once this is done, a standard Kalman filter program can be used to estimate the state vector, including the unknown parameters. The resulting algorithm is quite straightforward.

Consider the unknown parameter vector as part of the state vector and add the equation

$$\frac{\partial \theta}{\partial t} = \boldsymbol{w}_{\theta}(t) \tag{4-2}$$

where $\boldsymbol{w}_{\theta}(t)$ is a random vector. If the parameters to be estimated are constant, then $\boldsymbol{w}_{\theta}(t) = \boldsymbol{0}$.

Let $\boldsymbol{y}$ be a new extended state vector:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{x} \\ \theta \end{bmatrix} \tag{4-3}$$

equation (4-1) and (4-2) can be expressed as:

$$\frac{\partial \boldsymbol{y}}{\partial t} = f(\,\boldsymbol{y}, \boldsymbol{u}, t) + \boldsymbol{M}_y \boldsymbol{w}_y(t) \tag{4-4}$$

where

$$f(\,\boldsymbol{y}(t), \boldsymbol{u}(t), t) = \begin{bmatrix} G(\,\boldsymbol{x}, \boldsymbol{u}, \theta, t) \\ \theta \end{bmatrix} \tag{4-5}$$

$$\boldsymbol{w}_y(t) = \begin{bmatrix} \boldsymbol{w}(t) \\ \boldsymbol{w}_{\theta}(t) \end{bmatrix} \tag{4-6}$$

$$\boldsymbol{M}_y = \begin{bmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{4-7}$$

where $\mathbf{I}$ is an identity matrix. In addition we assume the following:

$$\mathrm{E}\{\, \mathbf{y}(0) \,\} = \mathbf{y}_0 \tag{4-8}$$

$$\mathrm{E}\{\, [\mathbf{y}(0) - \mathbf{y}_0\,]\,[\mathbf{y}(0) - \mathbf{y}_0\,]' \,\} = \mathbf{P}_0 \tag{4-9}$$

$$\mathrm{E}\{\, \mathbf{w}_y(t) \,\} = \mathbf{0} \tag{4-10}$$

$$\mathrm{E}\{\, \mathbf{w}_y(t)\, \mathbf{w}_y(\tau)' \,\} = \mathbf{Q}\ \delta(t-\tau) \tag{4-11}$$

where

$$\delta(t-\tau) = \begin{cases} 1 & \text{if } t = \tau \\ 0 & \text{if } t \neq \tau \end{cases}$$

and $\mathbf{Q}$ is the correlation matrix of the process noise.

Observations of the process are made at discrete times i = 1, 2, . . . N. The observations are a non-linear function of the state vector plus a random error component:

$$\mathbf{z}(t_i) = h\,[\mathbf{y}(t_i)\,] + \mathbf{v}(t_i) \tag{4-12}$$

where $\mathbf{z}$ is the measured output vector, $\mathbf{v}$ is the measurement noise vector, and

$$\mathrm{E}\,\{\, \mathbf{v}(t_i) \,\} = \mathbf{0} \tag{4-13}$$

$$\mathrm{E}\,\{\, \mathbf{v}(t_i)\ \ \mathbf{v}(t_j) \,\} = \mathbf{R}\ \delta_{ij} \tag{4-14}$$

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

It is desired to construct a "best" estimate of $\mathbf{y}(t_i)$ given past values of the measurement vector $\mathbf{Z}_I=\{\mathbf{z}(t_1), \mathbf{z}(t_2), \ldots, \mathbf{z}(t_i)\}$. The best estimate is the conditional mean of $\mathbf{y}(t_i)$ given $\mathbf{Z}_I$ That is

$$\hat{\mathbf{y}}(t_i \,|\, I) \equiv \mathrm{E}\{\, \mathbf{y}(t_i) \,|\, \mathbf{Z}_I \,\} \tag{4-15}$$

The EKF (C -D) yields a sequential algorithm that can be used to generate the conditional mean of $\mathbf{y}(t_i \,|\, I)$ and the mean of the conditional covariance matrix $\mathbf{P}(t_i \,|\, I)$.

$$\mathbf{P}(t_i \,|\, I\,) \equiv \mathrm{cov}\,\{\, \mathbf{y}(t_i)\ \mathbf{y}(t_i) \,|\, \mathbf{Z}_I \,\} \tag{4- 16}$$

Referring to Lewis (1986), we find that the best estimate for updating the state between observations is

$$\frac{\partial}{\partial t}\hat{y}(t\,|\,I\text{-}1)) = f(\hat{y}(t\,|\,I\text{-}1)), \boldsymbol{u}(t), t) \tag{4-17}$$

the error covariance matrix is update by

$$\frac{\partial}{\partial t}\boldsymbol{P}(t\,|\,I\text{-}1)) = \boldsymbol{F}(t)\,\boldsymbol{P}(t\,|\,I\text{-}1)) + \boldsymbol{P}(t\,|\,I\text{-}1))\boldsymbol{F}^{\mathrm{T}}(t) + \boldsymbol{M}_y(t)\boldsymbol{Q}\boldsymbol{M}_y{}^{\mathrm{T}}(t) \tag{4-18}$$

$$t_{i\text{-}1} \le t \le t_i$$

where $\boldsymbol{F}(t)$ is a matrix obtained by linearizing $f(y(t_i\,|\,t_{i\text{-}1}), \boldsymbol{u}(t), t)$ about the best current estimate

$$\mathrm{F(t)} \equiv \left.\frac{\partial f(\,y(t),\boldsymbol{u}(t\,),\,t\,)}{\partial y}\right|_{y=\hat{y}(\mathrm{t})} \tag{4-19}$$

After each observation, the estimate of the state is updated according to the following equations:

$$\hat{y}(t_i\,|\,I) = \hat{y}(t_i\,|\,I\text{-}1) + K_i[\boldsymbol{z}_i - h(\,\hat{y}(t_i\,|\,I)\,)] \tag{4-20}$$

$$\boldsymbol{K}_i = \boldsymbol{P}(t_i\,|\,I\text{-}1)\,\boldsymbol{H}^{\mathrm{T}}(\hat{y}(t_i\,|\,I\text{-}1)\,)\,[\boldsymbol{H}(\hat{y}(t_i\,|\,I\text{-}1)\,)\boldsymbol{P}(t_i\,|\,I\text{-}1)\,\boldsymbol{H}^{\mathrm{T}}(\hat{y}(t_i\,|\,I\text{-}1)\,) + \boldsymbol{R}\,]^{-1} \tag{4-21}$$

$$\boldsymbol{P}(t_i\,|\,I) = [\,\boldsymbol{I} - \boldsymbol{K}_i\,\boldsymbol{H}(\hat{y}(t_i\,|\,I\text{-}1)\,)]\,\boldsymbol{P}(t_i\,|\,I\text{-}1) \tag{4-22}$$

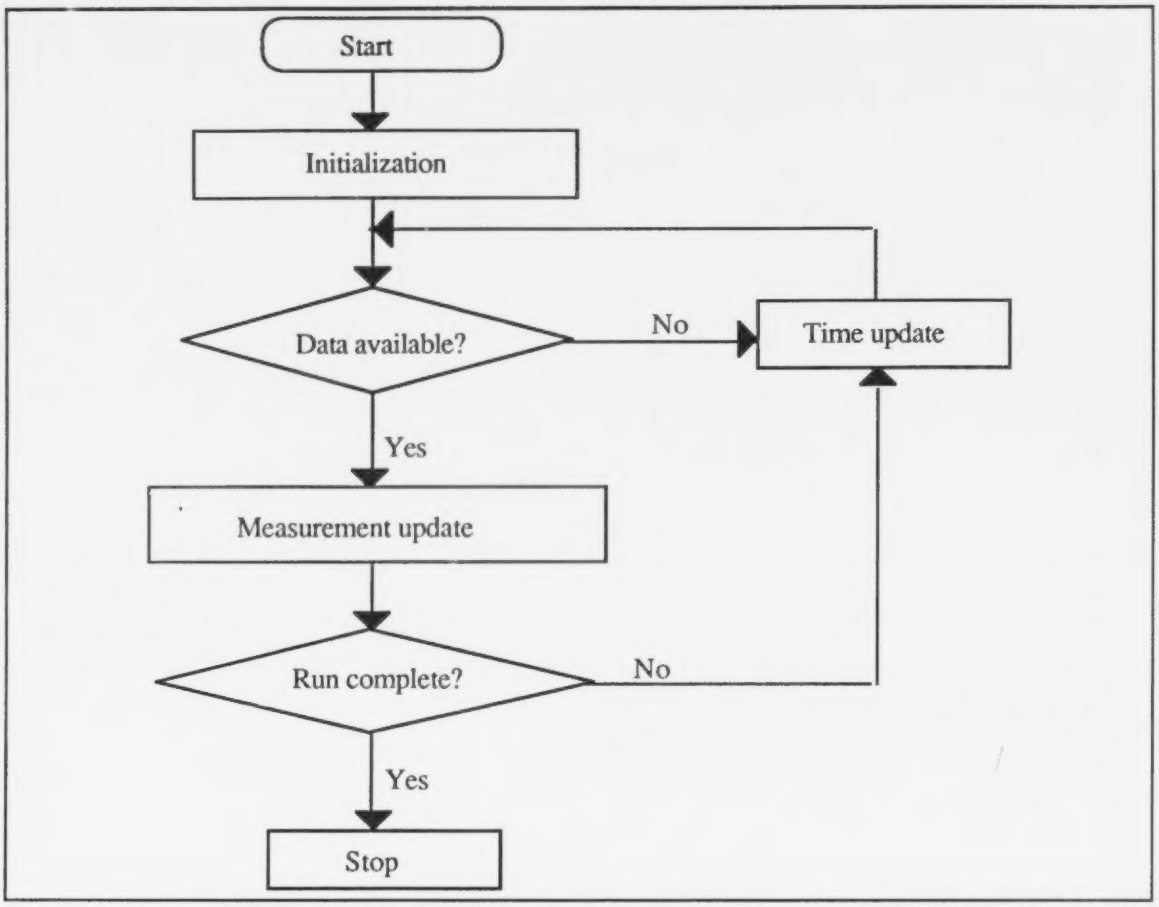
$$\boldsymbol{H}() \equiv \frac{\partial h\,(\,\boldsymbol{y}(t))}{\partial y} \tag{4-23}$$

Given initial values for $\hat{y}$ and $\hat{\boldsymbol{P}}$, equations (4-20) through (4-22) can be used to construct new estimates of the state—including the unknown parameters—following each observation.

The procedure has the following steps (Figure 4.1):

1. When $i = 0$, the filter is initialized by $\hat{y}_0 = [\,\hat{x}_0\ \hat{\theta}_0]'$ and $\hat{\boldsymbol{P}}_0$.
2. Time update: The prediction equations (4- 17) and (4-18) are integrated over time to produce $\hat{y}(t\,|\,I\text{-}1)$ and $\boldsymbol{P}(t\,|\,t_{iI})$
3. Measurement update: When a new observation is received, the update equations (4-20) and (4-22) are used to calculate new estimates $\hat{y}(t_i\,|\,I)$ and $\boldsymbol{P}(t_i\,|\,I)$. The gain matrix $\boldsymbol{K}_i$, is calculated using $\hat{y}(t_i\,|\,I\text{-}1)$
4. The process is then iterated.

Figure 4.1: Kalman Filter Estimation for Continuous State Equation and Discrete Observations


Start
Initialization
Data available?
No
Time update
Yes
Measurement update
Run complete?
No
Yes
Stop

5. Estimation of A Diffusion Model with Network Externalities by an EKF (C-D)

5.1 Designing the Kalman Filter

We are now ready to return to our original problem: how to estimate the demand and diffusion parameters for a diffusion model of a market with network externalities. Our first task is to transform the equations of section 2 into the state and observation equations of section 4. In our model, based on optimal dynamic pricing, the price p can be considered as the control variable. Our objective function is to find $p(t)$ which maximizes the present worth of future profits:

$$\text{Max} \int_0^\infty e^{-\delta t}\, [p(t)x(t) - c(x)]\, dt \tag{5-1}$$

subject to the constraints:

$$\frac{\partial x}{\partial t} = G\,(d^{\alpha}\,(x,p),x), \qquad x(0) = x_0 \tag{5-2}$$

$$0 \leq p \leq W(\,x, x\,) \tag{5-3}$$

where, x is the number of adopters, $c(x\,)$ is the monopolist's cost function and δ is the discount rate.

Dhebar and Oren (1985) show how to solve for the optimal price trajectory $p(t)$. They also show how to calculate the equilibrium values of network penetration, x^*, and price, p^* as the unique solution to the two equations:

$$p^* = W(x^*, x^*) \tag{5-4}$$

and

$$p^* - c_x^* - \frac{x^*}{G_d\,(x^*,x^*)d_p^{\alpha}}\,[G_d(x^*,x^*)\,d_x^{\alpha} + G_x(x^*,x^*) - \delta] = 0 \tag{5-5}$$

In particular, they show that the optimal p depends only upon x, and they derive the following differential equation for dp/dx, which can then be integrated backward from (x^*,p^*) to derive an optimal pricing rule.

$$\frac{dp}{dx} = \frac{d_p^{\alpha}}{x\,(\,G_{dd}\,(d_p^{\alpha}\,)^2 + G_d\,d_{pp}^{\alpha}\,)}\;[\,G_d\,(1 - x\,\frac{d_{px}^{\alpha}}{d_p^{\alpha}}\,) - x\,(\,G_{dd}\,d_x^{\alpha} + G_{dx}\,) -$$

$$- \frac{(\,p - c_x\,)\,G_d^2\,d_p^{\alpha} - x\,G_d\,(\,G_d\,d_x^{\alpha} + G_x - \delta)}{G}\;] \tag{5-6}$$

where subscripts denote partial derivatives.

In terms of the Extended Kalman Filter model described in section 4, $x(t)$ is our state variable, and $p(t)$ is the control variable. We thus have the following continuous Kalman filter model:

$$\frac{\partial x}{\partial t} = G\,(d^{\alpha}\,(x,p),x) + w(\mathrm{t}) \tag{5-7}$$

where w(t) is a random noise variable.

Since the number of adopters at time t_i is simply the number of subscribers to the monopolist's service, the observation equation is quite simple:[2]

---

[2] x has been normalized to [0,1]. In practice, one observes $N(t) = mx(t)$, where m is the total population.

$$z(t_i) = x(t_i) + v(t_i) \qquad (5\text{-}8)$$

where $v(t)$ is a gaussian random variable representing measurement error.

5.2 An Example Specification

In order to have an equation we can estimate, we have to specify more precisely our assumptions about G and $d^{\alpha}(x,p)$, and about $W(x,x)$. We consider the same case analyzed by Dhebar and Oren (1985). The willingness to pay function $W(\eta,x)$, is given by equation (2-5). Thus, the consumer's reservation prices decrease linearly with consumer index and increase linearly with the number of adopters. The willingness-to-pay of a marginal subscriber, $W(x,x)$ is given by the quadratic function:

$$W(x,x) = W_0(1-x)x \qquad (5\text{-}9)$$

The network's growth rate is assumed to be linear in the unsatisfied subscription demand and is given by equation (2-6). Finally, the vendor's cost function is given by the linear function in (2-7), and the parameter C_0 is assumed to be known to the vendor. We also fix the value of the anticipation parameter α at $\alpha=1$.

This then gives us the following expression for $d^{\alpha}(x,p)$

$$d^{\alpha}(x,p) = \frac{1}{2}\left[1 + \left(1 - \frac{4p}{W_0}\right)^{\frac{1}{2}}\right] \qquad (5\text{-}10)$$

which when substituted into (2-6) gives the following state equation:

$$\frac{\partial x}{\partial t} = b\left\{\frac{1}{2}\left[1 + \left(1 - \frac{4p}{W_0}\right)^{\frac{1}{2}}\right] - x\right\} \qquad (5\text{-}11)$$

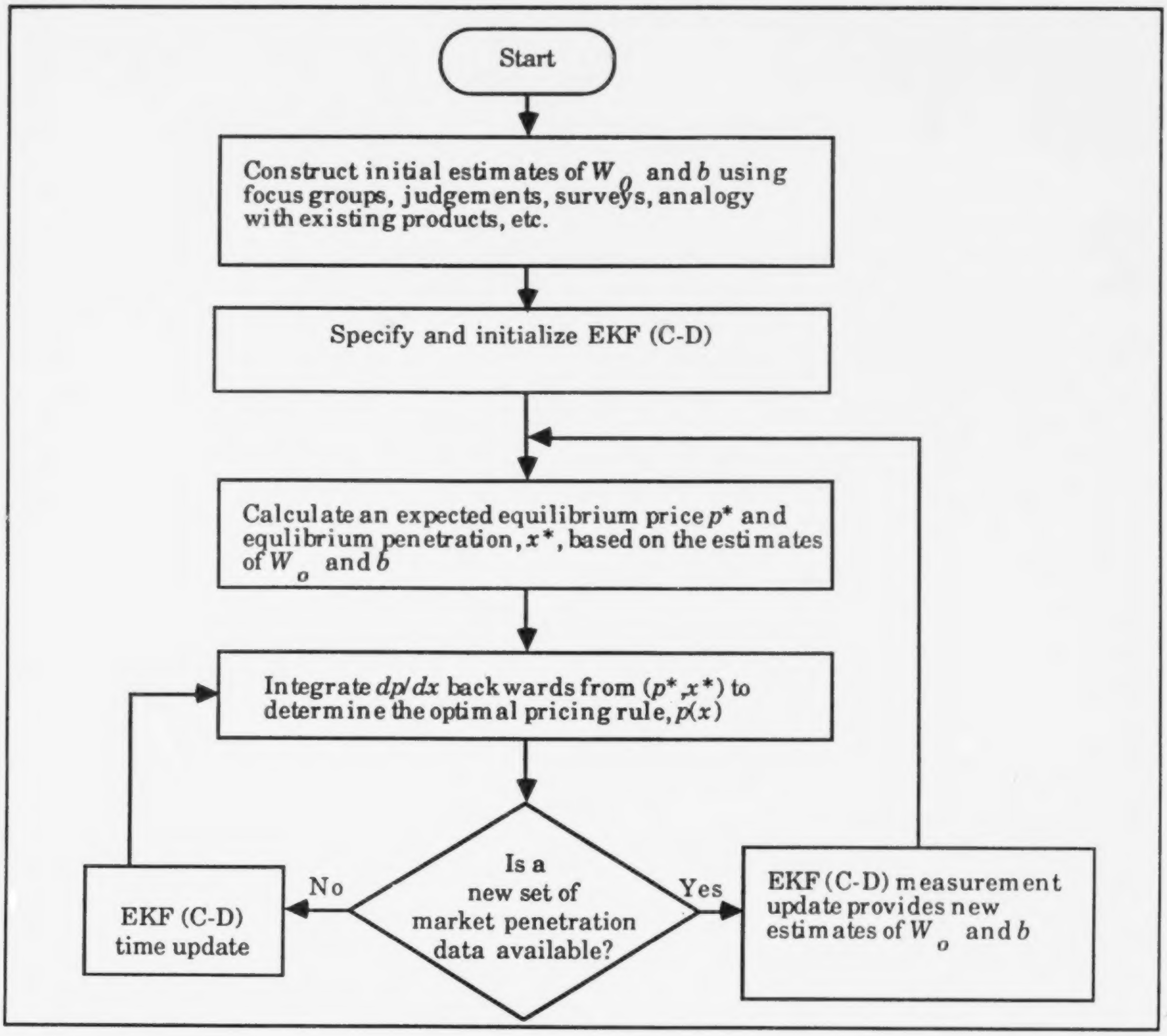
The Extended Kalman Filter state variable then has the form:

$$y = \begin{bmatrix} x \\ W_0 \\ b \end{bmatrix} \qquad (5\text{-}12)$$

6. Linking Parameter Estimation and Optimal Pricing

Dhebar and Oren have shown how to solve for the optimal price trajectory when W_0 and b are known. However, especially for a new product or service, it is unlikely that these parameters will be known with any great degree of accuracy. Rather than develop a pricing rule $p(x)$ in advance, and sticking to it without reference to the marketplace reaction, an intelligent vendor will observe $x(t)$ in the early stages of the process, compare it to the anticipated value, and react accordingly to alter his pricing policy. The use of an Extended Kalman Filter model to estimate the parameters $\theta=(b, W_0)$ provides the optimal linear estimates of these parameters for us in revising the pricing policy. This leads to the following strategy for the vendor (Figure 6.1).

Figure 6.1: Procedure for parameter estimation and update of optimal pricing


Start
Construct initial estimates of W_o and b using focus groups, judgements, surveys, analogy with existing products, etc.
Specify and initialize EKF (C-D)
Calculate an expected equilibrium price p^* and equlibrium penetration, x^*, based on the estimates of W_o and b
Integrate dp/dx backwards from (p^*, x^*) to determine the optimal pricing rule, $p(x)$
Is a new set of market penetration data available?
No
EKF (C-D) time update
Yes
EKF (C-D) measurement update provides new estimates of W_o and b

Initialization

1. Using focus groups, judgement, surveys, analogy with existing products, etc. construct initial estimates of W_O and b.
2. Set up an EKF (C-D) model for estimating the values $x(t)$, W_O and b. For initial values, use the estimates from step 1, and $x(0) = 0$.

Iteration

3. Based on the estimates of the parameters W_o and b, calculate an expected equilibrium price p^*, and equilibrium penetration x^*.
4. Integrate dp/dx backwards from (p^*,x^*) to determine the optimal pricing rule $p(x)$.
5. If an observation of $x(t)$ is available, go to step 6. Otherwise, integrate the estimate $\hat{x}(t)$ forward using the state equation (5-10) and the pricing rule $p(x)$ calculated in step 4.
6. Periodically observe the actual level of adoption. Using the EKF, construct new estimates of W_o and b.
7. Repeat from Step 3.

In plain terms, after each observation of the state of the market, new estimates of the key parameters are used to develop a new pricing rule which is then followed until the next update.

Note that, at each observation interval we calculate an entire new pricing policy, $p(x)$. This policy itself implies that the vendor is continuously changing prices as x, the degree of market penetration, varies. We are not calculating a single price at each observation, which might then be held fixed until the next update. Rather, we are calculating a complete pricing rule—which prescribes continuously varying prices—to be applied until a new, and presumably better rule can be calculated.

The procedure developed here is very general. It can be used not only for a market with a network externality, but for any market where the optimal pricing policy depends only upon the current system state. Moreover, as our simulation results show, the method produces a result which is very insensitive to errors in the original parameter estimates, achieving near optimal results even when the initial values are off by 50%.

7. Testing the EKF (C-D) Estimation Procedure in a Simulated Market.

7.1 The Simulation Procedure

In order to test the efficacy of the procedure described here, we simulated a market characterized by a demand curve as in equation (2-5) and a diffusion process as described by equation (2-6) by integrating the diffusion equation for a range of parameter values and pricing policies. For our base case, we selected parameters as follows:

$$\begin{aligned} W_o &= 1 \\ C_o &= .05 \\ b &= 1 \\ \delta &= .1 \end{aligned} \tag{7-1}$$

We assumed the process noise $w(t) = 0$. We assumed that the observation errors $v(t)$ were normally distributed with mean 0 and variance equal to $0.05x$. Observations and measurement updates were presumed available at intervals of $\Delta\tau = 0.01$ time unit.

Using equations (5-4) and (5-5), we calculated, for the base case, the equilibrium point (p^*,x^*). Equation (5-6) can be integrated backwards from (p^*,x^*) to derive the optimal pricing rule $p(x)$. The time to equilibrium, t^*, assuming an optimal penetration pricing strategy was followed, can be calculated from:

$$t^* = \sum_{i=1}^{n} \Delta t_i = \sum_{i=1}^{n} \frac{(x_i - x_{i-1})}{\left.\frac{\partial x}{\partial t}\right|_{t_i}} = \sum_{i=1}^{n} \frac{(x_i - x_{i-1})}{b(d^{\alpha}(x_i, p_i) - x_i)} \tag{7-2}$$

where $x_n = x^*$. The first line of Table 7.1 shows these values for the base case.

Using the pricing policy thus generated, we simulated the evolution of the market by integrating the diffusion equation, using the Runge-Kutta technique, over the interval $t \in [0, t^*]$ using the optimal pricing rule $p(x)$ calculated previously. We verified that the process indeed converged at the analytically derived equilibrium (p^*, x^*) at time t^*. We used the values of $x(t)$ from the simulation to calculate Π_1, the present worth of net profits earned, assuming that there is no further evolution of the market once equilibrium is reached. Define Π^* as the present worth of net profits for the base case.

Next we asked, by how much would the vendor mis-estimate the market equilibrium if he used different values for W_o and b in making his forecast. Lines 2-5 of Table 7.1 show how the vendor's estimates of $(\hat{p}^*, \hat{x}^*)$, and $\hat{t}^*$ would change if his estimates of the parameters were too high or too low. If the vendor believed his values of W_o and b to be correct, he would calculate his *forecast* profits Π_1, exactly as we did in the previous step, arriving at either an over- or under-estimate of the profits attainable, depending upon the errors in W_o and b. These profit forecasts are also show in Table 7.1.

Table 7.1: Forecast Profits

Initial parameter	$\hat{x}^*$	$\hat{p}^*$	t^*	forecast profit, Π_1
Base case	.6315	.2327	6.077	1.0603
W_o = -20%	.6250	.1875	6.114	.7898
W_o = -50%	.6046	.1195	6.235	.386
b = -20%	.6294	.2333	7.61	1.0376
b = +20%	.6329	.2323	5.058	1.076
b = - 50%	.6325	.2348	12.24	.9837
b= + 50%	.6344	.2319	4.041	1.0904
W_o = -20% and b= -20%	.6230	.1890	7.656	.7767

We then asked, how would the vendor actually fare if it blindly followed the pricing rule developed on the basis of initial estimates of the parameters, and held to the rule even if the observed results in the market were not exactly as predicted. To answer this question, we integrated the state equation forward using the "true" —*e.g.* the base case—values of the parameters, but at the same time using the pricing rule calculated on the basis of faulty a priori estimates of these parameters. The equilibrium time t^* was defined to be the time when $dx/dt \leq .002$. x^* and p^* are the values reached at time t^*. This produced the results Π_2, in Table 7.2.

Table 7.2: Constant Pricing Rule

Initial parameter estimates	x^*	p^*	t^*	profit , Π_2
Base case	.6315	.2327	6.077	1.0603
W_o = -20%	.7496	.1875	7.38	.9381
W_o = -50%	.8608	.1195	7.6301	.5424
b = -20%	.6290	.2332	6.14	1.0600
b = +20%	.6326	.2323	6.13	1.0602
b= -50%	.6231	.2347	6.1501	1.0597
b= +50%	.6341	.2319	6.1201	1.0602
W_o = -20% and b= -20%	.7488	.1879	7.3801	.9399

Finally, we integrated the state equation forward assuming that the vendor started off with a non-optimal pricing strategy due to faulty prior estimates of the demand and diffusion parameters, but used an EKF (C-D) model to update those initial estimates, and revised his pricing rule accordingly. Again, the simulation proceeds using the "true" values of W_o and b, but the pricing rule is gradually aligned with the correct rule as the vendor improves his estimates of the correct parameter values and recalculates his pricing rule accordingly. Table 7.3 shows the the final values of x^*, p^*, t^* and the value of profits, Π_3, actually realized. It also compares Π_2 and Π_3 with Π^*, the maximum profits attainable with perfect prior information. Note that in several cases, Π_2 and Π_3 exceed the theoretical maximum Π^* by amounts of less that 0.2%; we attribute this to inaccuracies resulting from our approximate method of determining t^*.

Table 7.3: Parameters and Pricing Rule Updated by EKF (C-D) Estimation

Initial	$\hat{x}^*$	$\hat{p}^*$	x^*	p^*	t^*	Π_3	$\frac{\Pi_2}{\Pi^*}$	$\frac{\Pi_3}{\Pi^*}$
W_o = -20%	.625	.1875	.6303	.2329	6.89	1.0593	.884	.9991
W_o = -50%	.6046	.1195	.6305	.2329	6.82	1.0600	.511	.9997
b = -20%	.6294	.2333	.6305	.2329	6.82	1.0601	.9997	.9998
b = +20%	.6329	.2323	.6305	.2329	6.59	1.0626	.9999	1.002
b = -50%	.6325	.2348	.631	.2327	6.78	1.0612	.9994	1.0008
b = +50%	.6344	.2319	.6310	.2327	6.88	1.0596	1.001	.9993
W_o = -20% and b= -20%	.623	.189	.6303	.2329	6.89	1.0593	.886	.9993

7.2 Discussion of Results

Just as we would expect, Table 7.1 shows that the the use of incorrect estimates of the parameters produces forecasts of future profits which can be wide of the mark. For example, if the value of W_0 is underestimated by 50%—i.e. the demand is underestimated—then the vendor will forecast 50% less profit than he would actually realize if he correctly estimated the demand. An underestimate of W_0 leads the vendor to price too low. Left uncorrected (*cf* Table 7.2), the resultant pricing strategy leads to an equilibrium at a higher value of x^*, but a lower price, p^*, with consequently lower profits. When the EKF is used to revise the estimates of W_0 and b as sales reports come in, there is significant improvement. Within a time period of 1.9, the EKF has produced estimates which are within 1% of the true values of the parameters W_0 and b. The result is a realized profit which differs by less than 0.03% from what could have been achieved with perfect information.

By contrast, an error in the initial estimate of the rate of diffusion has a much smaller influence on profits realized. Consider the case where b is initially estimated at only 50% of its actual value. The vendor will expect the time to equilibrium to be nearly twice as long and the profits to be 12% less than if the correct values had been used. But the equilibrium values x^* and p^* will not be forecast to be very different. Moreover, for the range of values we have assumed here, the pricing rule which will be followed will not differ very much from the rule which would have been followed if b had been known exactly (*cf* equation 5.6). The result (Table 7.2) is that the profit realized even when following a non-optimal rule is not terribly different from what could be realized from a better estimate of b. For example, if pricing is based on an estimate of b which is 50% too low, actual profits realized will be less than 0.06% below what would have been achieved with perfect information (Table 7.2). Still, the EKF is able to ascertain the correct values of W_0 and b within less than one time unit.

We also considered the case where the prior estimates of W_0 are too high. This leads to a forecast equilibrium (and associated pricing rule) which is above the demand curve. What happens when the pricing rule calls for prices in excess of the williness of any consumer to pay i.e. $p > W_0/4$? At that point, the "unsatisfied demand" should be negative, and the market would begin to decline, thus failing to reach critical mass and a stable equilibrium.

In particular, since the optimal pricing rule is monotonic in p, then if the initial estimate of $p^* > W_0/4$, the pricing rule must eventually set price above the demand curve such that there is no solution to

$$d^{\alpha}(x,p) = \max\{\eta:\ 0 \le \eta \le 1,\ W(\eta, [\alpha\eta + (1-\alpha)x]) = p\}$$

This leads to undefined behavior in our model, though in practice it must certainly lead to a negative rate of growth. Further work is necessary to properly specify the behavior of the system under these conditions, particularly in the EKF case where the observation of negative growth can be used to correct the estimates which led to the inappropriate pricing rule in the first place.

For the base case values we have chosen, an estimate of W_0 which is more than 8% too large results in $p^* > W_0/4$, and thus undefined behavior.

8. Conclusion

Penetration pricing is an attractive strategy for accelerating the expansion of a market with network externalities. With perfect knowledge of the demand curve and the diffusion equation, it is possible to calculate a pricing policy which maximizes the present value of future profits. However, the paramters of the demand and diffusion processes are unknown a priori. We have shown that a continuous differential equation Extended Kalman Filter with discrete observations (EKF (C-D)) model can be used to estimate the demand and diffusion parameters from observations of early market data. Thus errors in a priori estimates of these parameters can be corrected and the optimal pricing policy realized. EKF (C-D) is a particularly attractive estimation method, because one can begin updating the estimates with the very first observations of the market, unlike OLS techniques which would require a significant number of observations before the first parameter estimates would be available. Simulation of our model shows that achieved results are far more sensitive to errors in the estimates of the demand curve than of the diffusion equation. It also shows that for errors of 20-50% in the a priori estimates of the parameters the model converges to the correct values sufficiently rapidly to produce profits which differ typically by less than 0.1% from what would have been realized with perfect a priori information.

The current formulation of our model breaks down if demand is overestimated to the point that the vendor is led to price totally above the curve $W_0(x,x)$. Further work is needed to extend the model to properly treat initial errors in the estimates of this magnitude.

The Extended Kalman filter technique is a very general one with applicability for estimating parameters of diffusion processes other than those considered here. Further exploration of the properties of EKF estimators for these types of applications seems warrented.[3]

References

Artle, R. and C. Averous (1973), "The Telephone System as a Public Good: Static and Dynamic Aspects," *Bell Journal of Economics and Management Science*, 4 (spring), 89-100.

Bass, F. M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15, January, 215-227.

Dhebar, Anirudh and Shmuel S. Oren (1985), "Optimal Dynamic Pricing for Expanding Networks," *Marketing Science* Vol. 4, Fall.

Kahl, Douglas R. and Johannes Ledolter (1983), "A Recursive Kalman Filter Forecasting Approach," *Management Science*, Vol. 29, No. 11, November.

Makridakis, S. and S. C. Wheelwriht (1977), "Adaptive Filtering: an integrated autoregressive/moving averege filter for time series forecasing," *Opl Res. Q.* 28, 425-437.

Makridakis, S. and S. C. Wheelwriht (1978), *Forecasting Methods and Applications* , Wiley, New York.

Rabenau, B. and K. Stahl (1974). "Dynamic Aspects of Public Goods: A Further Analysis of the Telephone System," *Bell Journal of Economics and Management Science*, 5 (Autumn), 651-669.

Rogers, E. M. (1983), *Diffusion of Innovations*, 3rd ed. New York: The Free Press.

Rohlfs, J. (1974), "A Theory of Interdependent Demand for a Communication Service," *Bell Journal of Economics and Management Science*, 5 (spring), 16-37.

Schmittlein, D. and V. Mahajan (1982), "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, 1 (Winter): 57-78.

---

[3] The authors wish to acknowledge many helpful discussions with Samuel Anderson and Steven Garber.

RENEGING BY CABLE OPERATORS ON THEIR FRANCHISE PROMISES: OPPORTUNISM OR ECONOMIC NECESSITY?

by

Mark A. Zupan*

Department of Finance and Business Economics
USC School of Business
Los Angeles, CA 90089-1421
(213) 743-5778

Revised, May 1988

*The author would like to thank Richard Craswell, Doug Joines, Paul Joskow, Ken Koford, Michael Levine, Tony Marino, Herman Quirmbach, Richard Schmalensee and Clas Wihlborg and workshop participants at the University of Arizona, USC and UCLA for helpful comments and suggestions. Elden Chang, David Franzblau, Chris Murdoch, William Stevens, and Bob Townsend have provided research assistance.

I. INTRODUCTION

Through a seminal case study, Oliver Williamson (1976) highlights some of the central problems associated with ex ante franchise bidding competitions as a mechanism for controlling ex post natural monopolies. Among the noted problems is that winning bidders can renege on their franchise commitments and may do so for opportunistic reasons. The possibility of opportunistic behavior stems from the idiosyncratic investments associated with a franchise contract. After the specific investments are made, quasi-rents are created (cf., Benjamin Klein, Robert Crawford, and Armen Alchian, 1978). These quasi-rents, or differences between an asset's value in its current use and its value in its next best use, provide bidders with the temptation to over-promise during the bidding process. If a misleading bid gets the bidder's foot in the door, franchising authorities can be "held up" (Victor Goldberg, 1976) for contractual concessions.

Based on Williamson's microanalytic analysis of the city of Oakland's cable franchising experience, one may be tempted to infer that opportunism plagues CATV franchising bidding processes. The winning bidder in Oakland promised to charge a price over 50 percent lower than the next lowest bidder. Once operations began, however, the winner negotiated favorable contractual changes: the construction deadline was extended; the channel capacity was downgraded; and prices for additional outlets were raised.

Franchise case histories paralleling that of Oakland's are not difficult to come by. Indeed, the renegings, primarily in certain large cities, that have been extensively reported in the popular press all appear to support Williamson's hypothesis that opportunism is an unavoidable and unfortunate fact of cable franchising life. Four years into the Denver contract, for example, Mile Hi Cablevision eliminated or postponed plans for $32 million in services that it had originally promised to provide.[1] After spending $160 million wiring Dallas, Warner-Amex secured concessions worth $20 million in capital costs and $5 million annually in operating costs.[2] Three years after cabling New Orleans, Cox Cable obtained a 36 percent basic rate increase and significant scalebacks in its community programming commitments.[3]

While the preceding examples illustrate the enforcement problems that may be associated with franchise contracts, they are deficient in two respects. First, individual case studies do not provide systematic evidence on the extensiveness of promise-breaking by cable operators. Williamson (1976, p. 91) indeed makes clear that the Oakland case study cannot claim to be representative of the franchising experience; rather, the careful examination of a single data point is intended to reveal that many of his concerns with franchise bidding "are not purely imaginary."

Second, renegotiations need not imply guileful behavior on the part of operators. Reneging operators, that is, may not be distorting the information available to them. The current information may simply be less optimistic than was initially estimated: costs may be higher and/or demand may be lower than projected. And it may be wise, in both the efficiency sense and in terms of the long-run interests of both parties, to incorporate the new information into the franchise contract and to move on from there (Klein and Roy Kenney, 1986; and Timothy Muris, 1981).

The justifications operators offer when requesting scalebacks are typically cloaked in pronouncements that the unforeseen has occurred. For example, Oakland's operator argued that construction was proceeding slower than anticipated and that fewer households were subscribing than originally estimated (Williamson, 1976, p. 96). Warner-Amex requested concessions in

Dallas claiming an unexpectedly low market penetration rate. Warner-Amex pointed to difficulties in getting residents of multiple dwelling units (MDUs) to subscribe in the face of fierce competition from satellite master antenna television (essentially a miniature cable system located in a MDU or hotel).[4] In Boston, Cablevision's requests for relief were predicated on an overestimation of market size, difficulties in wiring apartments and property taxes nearly double what the firm had anticipated.[5]

Confronted by the rationales offered by reneging operators, one might be inclined to discount or even dismiss their validity. Still, operators may have strong reasons not to engage in opportunistic promise-breaking. Among these are reputational constraints (Klein and Keith Leffler, 1981) and the potential for vertical integration by cities into the distribution of CATV services (Paul Joskow, 1985; and Klein, Crawford, and Alchian, 1978). The ability of operators to exploit the advantages of incumbency is also circumscribed by the countervailing power of cities to do likewise; while their irreversible investments place existing firms in a superior position vis-a-vis potential rivals, the sunk costs have value only to the city in which they are sunk.

This paper addresses the deficiencies in the existing literature pertaining to opportunism and the efficacy of franchise bidding schemes. First, to ascertain how well Williamson's microanalytic observations apply to cable markets in general, systematic information is provided on the extent of reneging by operators. Second, operators' market misjudgments are examined to determine whether they evidence pervasive operator opportunism or unforeseeable changes in market information and generally well-working restraints on operator opportunism.

II. THE DETERMINANTS OF OPERATOR OPPORTUNISM

Operator opportunism is not a relevant problem when the opportunity cost of walking away from the franchise relationship is nonpositive for the city involved. If a city can costlessly extricate its assets, hold ups by operators will not be threatened and need not be feared. Only when a city's assets have values specific to an existing franchise relationship--values which are higher when the assets are internal to the relationship rather than external--does the potential for operator opportunism arise.

The reasons why a city's assets may be more valuable within an existing franchise relationship are manifold (cf., Joskow, 1985; and Williamson, 1976). These reasons may be classified into three general types:

1. franchise relationship search costs;
2. franchise relationship development and upkeep costs; and
3. franchise relationship exit costs.

Finding a suitable cable operator is generally not costless. By the time a franchise has been awarded, the expenditures associated with screening and selecting an appropriate operator have already been incurred and are bound to be at least partially irretrievable or sunk. Other things being equal, consequently, a city will have an incentive to stay with the franchise relationship into which it has already sunk search costs.

Once found and initiated, the typical cable franchise relationship demands nurturing and sustenance. Investments (e.g., time and effort) must be made in learning how to deal with the chosen cable firm and its management. Because the worth of such investments is likely to depreciate in value should the franchise be reassigned to a different operator, they provide an additional incentive to a city to retain its chosen operator.

Finally, the termination of a franchise relationship may impose substantial costs on a city and its policymakers. For example, the eviction of an existing cable operator may result in the loss of service between the time the old operation is shut down and the new system is built.[6] Any such termination is likely to produce a political outburst from constituents who lose cable service; an outburst which may make local policymakers more wary about evicting the existing operator.

II.A. The Role of Imperfect and/or Costly Information

Holding the size of a city's nonsalvageable assets constant, three factors may increase the incidence of operator opportunism:

1. costless, symmetric but nonetheless imperfect information;
2. perfect information that is equally available to both parties to a franchise relationship but costly to obtain or process; and
3. asymmetries in either the information sets or in the information-processing abilities of the parties to a franchise relationship.

Imperfect information makes it impossible to write complete contingent claims contracts; contracts specifying an operator's duties in every attainable state of nature. If all possible states cannot be foreseen, franchise contracts will be incomplete and there will be room for cable operators to behave badly *ex post* (Joskow, 1985).

Provided that all potential states of nature can be foreseen, the writing of complete contracts may nonetheless be impracticable if cities find the available information costly to acquire or process (cf., Roy Radner, 1968, and Williamson 1985). Some of the hindrance may be due to bounded rationality (Herbert Simon, 1959, 1961 and 1978); i.e., the mind itself may be a scarce resource requiring economization in the face of the multitude of bits of information that it confronts.

Last, asymmetries in the information sets or in the information-gathering abilities of the parties to a franchise relationship can leave room for operator opportunism. Adverse selection is perhaps the most oft-cited example of bad behavior arising from asymmetric information (George Akerlof, 1970). If cities cannot distinguish good operators from "lemons" and if the sellers of lemons are unwilling to candidly reveal the true quality of their product, the lemons may drive good operators out of franchise relationships altogether.

III. NONOPPORTUNISTIC PROMISE BREAKING BY OPERATORS

While imperfect information may foster promise breaking by operators, the reneging which follows the arrival of new information need not represent opportunism. It may be both fair and efficient to revise contracts in response to new information (cf., Richard Posner, 1986). Klein and Kenney (1986, p. 2) argue that contracts are kept purposefully flexible:

> Since most unlikely events can be accommodated to at lower cost after the relevant information is revealed, expenditures made by transactors to anticipate such events and to negotiate prespecified contractual responses are allocatively unnecessary.

Muris (1981, p. 533) elaborates on the nonopportunistic reasons for contract revision in the face of unexpected cost increases. He puts forth the example of a case which involved a contract to excavate a cellar:[7]

> Soon after excavation started, the excavator hit solid rock, and the parties then agreed that the excavator would receive a nine-fold price increase. Several reasons unrelated to opportunism may explain why the owner of the property agreed to pay a higher price. For instance, the owner might gain a reputation for fair dealing, avoid

> driving the excavator into bankruptcy, prevent the excavator from giving poorer-than-normal performance, or feel enhanced self-esteem from acting altruistically.[8]

IV. THE CONSTRAINTS ON OPERATOR OPPORTUNISM

While distinguishing nonopportunistic from opportunistic operator reneging may be exceedingly difficult, there are at least five important mechanisms which serve to restrain operators from breaking their promises for opportunistic reasons. These mechanisms include:

1. penalty provisions written into a contract coupled with any legal costs a reneging operator incurs when the penalty clauses are invoked;
2. vertical integration;
3. cross-sectional reputation--a badly behaving operator may be unable to obtain new franchises if an existing contractual city-partner is maltreated;
4. reputation internal to an existing franchise relationship--bad behavior by an operator may induce retaliation and lower-than-optimal investment in the relationship by the operator's contractual city-partner; and
5. monopsony power possessed by an operator's contractual city-partner--the irrevocable investments made by cable operators in the existing franchise relationship have value only to the city in which they are sunk and this value is thus subject to expropriation by the city.

IV.A. Penalty Clauses

The penalty provisions typically found in a cable franchise contract are "hostages" left by the operator in the hands of the city and/or third parties to ensure that the operator will uphold its side of the bargain. To the extent that a city can at least partially discern the reasons behind an operator's request for contractual revisions (Muris, 1981), such hostages are more likely to be executed in response to bad conduct. Penalty clauses thus provide a disincentive to opportunistic promise-breaking by operators.

IV.B. Vertical Integration

The potential for vertical integration by cities into local cable distribution likewise limits opportunistic promise-breaking by operators. Where prospective city-partners foresee that a franchise relationship will require the investment of nonsalvageable assets and that the quasi-rents associated with these assets may be ripe for appropriation, vertical integration may prove more desirable than the striking of a long-term, incomplete franchise contract (Klein, Crawford, and Alchian, 1978). Systematic empirical evidence supporting this hypothesis continues to amass. From the automobile industry (Kirk Monteverde and David Teece, 1982), the organization of production in the aerospace industry (Scott Masten, 1984), supply arrangements between coal producers and electric utilities (Joskow, 1985), and the provision of freight-rail services (Thomas Palay, 1984), vertical integration appears to be more prevalent where the risks of opportunistic behavior are greater.

In local cable markets, vertical integration takes the form of municipal ownership: i.e., the proprietorship and operation of the system by the city. Perhaps tellingly, less than one percent of all cable systems in the U.S. are municipally owned and only .1 percent of all subscribers are served by such systems.[9]

IV.C. Cross-Sectional Reputation

Reputational concerns may also serve to promote upstanding conduct by cable operators. Previous authors (cf., Klein and Leffler, 1981) have correctly pointed to the lost business that "notorious" actions may entail. Importantly, future business may be lost either from existing or from prospective contractual partners.

In the case of CATV franchising, the threat of lost business from prospective contractual partners (i.e., cross-sectional reputation) may be particularly effective at curbing opportunistic behavior by operators. Of the 36 million cable subscribers in 1985, well over 90 percent were accounted for by multiple system operators (MSOs).[10] To the extent that cities place stock in a cable firm's performance elsewhere and that an operator intends to win future franchise bidding competitions held by other cities and/or assure local policymaker approval of prospective system purchases, abstention from opportunistic behavior becomes desirable.[11]

Cities appear to place special emphasis on a bidder's track record when awarding franchises. To justify why bidders' applications were approved or denied, for example, Massachusetts communities invariably pointed to records of past performance:[12]

> Cable systems managed by this company [Pioneer] in adjacent areas show a great deal of dissatisfaction with past performance. (Northampton, 1979)
>
> It has come to the attention of the Board through information received from the Connecticut Department of Public Utility Control that a charge of misleading marketing practices was levied against Southern Connecticut Cablevision in 1978....The fact that charges were levied and corrective action ordered by the State greatly concerns the Board of Selectmen. Reliability and honesty will serve as the basis for a positive working relationship between the applicant and the Town of Plymouth, and the above information runs contrary to those standards. (Plymouth, 1980)
>
> The history of the applicant [Warner], even in the recent past, with regard to forfeiture, sale, surrender

> and revocation of cable television franchises in small cities and towns is unacceptable. (Gloucester, 1980)

> My [the mayor's] investigation of Continental's performance in building cable systems in other communities indicates that the company builds as promised. (Brockton, 1981)

> [Continental] Has never sold a system, never lost a license, nor failed to have a franchise license renewed..excellent responses from other communities to CAC questionnaire [mailed by the Cable Advisory Committee to other communities served by the applicants]... Weaknesses: Only fair responses to CAC questionnaire from Newton. (Needham, 1982)

> The reasons for the Issuing Authority's approval... include...Continental's expertise in constructing and operating cable systems, its record of meeting franchise commitments...its excellent reputation in the cable industry. (Marblehead, 1984)

IV.D. Reputation Internal to the Existing Relationship

In addition to being valued cross-sectionally, a reputation for virtue may be highly prized within an existing franchise relationship. Where the worth of the final product depends on both party's effort levels over the course of the relationship, brazenly self-serving conduct by an operator may provoke retaliatory treatment by the city and leave both sides with a lower-valued final product (Tracy Lewis, 1986, and Stewart Macauley, 1963). Indeed, there are many ways, besides invoking penalty clauses, that a city could make an operator's life unpleasant in response to an opportunistic hold up. The operator may be called before the city council more frequently. The city may increase property taxes, hike utility pole rental rates or drag its feet on granting easements. Dogged persistence may be displayed to ensure that the operator adheres to the fine print of the contract, no matter how small the fine print. The franchise contract may not be renewed when it expires.

Virtue proves valuable within an ongoing relationship for the same reason that the maximization of joint benefits is more likely when a rivalry is continuous (cf., Robert Axelrod, 1984; Thomas Schelling, 1960; and Frederic Scherer, 1980); and cooperation may emerge between competitors who, at any point in time, compete in numerous markets (Corwin Edwards, 1955; and Scherer, 1980). According to Scherer (1980, p. 163), "Repeated experience under stable conditions affords an opportunity for learning to cooperate and trust one another; and when the game will be played continuously or repeatedly, each party can threaten its rivals with damaging retaliation tomorrow if cooperation is not forthcoming today."[13]

The importance of internal reputation in cable markets is illustrated by the lengths companies will go to to maintain close ties with the communities they serve. The president of the MSO Colony Communications argues that:[14]

> when a franchise is won, that becomes the time for the MSO to become more intense in its communications with the community 'not only to promote itself, but more importantly to establish solid good will.'...[Colony] sends its director of governmental relations to each of its franchise cities on a regularly quarterly basis to give first-hand presentations to franchise authorities ...[the presentations] include...what new programming services Colony is adding or why they are taking off a service, how the community can best utilize its public access channels, why rate increases become necessary, and an interface with authorities on what needs the communities may have...'In essence, we are trying to head off complaints and difficulties at the pass.'

IV.E. Cities' Monopsony Power

Previous analyses (cf., Williamson, 1976, p. 83 and 89) argue that the irrevocable commitments made by operators "place winners of the original competition at a substantial advantage over nonwinners" and convert "what was once a large numbers bidding situation, at the time the original franchise was awarded...into what is tantamount to a small numbers bargaining situation" once the franchise relationship is initiated. What the prior studies have failed to emphasize, however, is that incumbent cable firms have in a sense earned these nontrivial advantages; resources have been expended by the firms to acquire the advantages of incumbency. The ability of operators to exploit the advantages of incumbency, furthermore, is constrained by the fact that an operator's nonsalvageable investments have value to only one buyer: the city in which the investments have been sunk. While prior research has typically pointed to the monopoly power opportunistic operators may exercise, the counteracting monopsony power cities may exert may be substantial. And cities may be as capable as operators of behaving badly.[15]

V. THE FREQUENCY OF AND POSSIBLE REASONS FOR OPERATOR RENEGING

The arguments of the preceding section suggest that cities possess a potent arsenal of defenses against any opportunistic forays by cable operators. Whether these enforcement mechanisms deter bad conduct is naturally an empirical question. Evidence from the single data point observed by Williamson (1976) intimates that the efficacy of available contractual safeguards is low. More systematic evidence, however, suggests otherwise. If anything, reneging by operators appears to be infrequent and motivated by reasons other than opportunism.

V.A. Evidence from Local Newspapers, the Trade Press, and Surveys of Local Policymakers and Cable Operators

Robin Prager (1986) examines the experiences of 92 Massachusetts communities which franchised cable systems over 1973-1981. By combing through the local newspaper clippings kept on file by the Massachusetts CATV

Commission, Prager concludes that in terms of the three important dimensions of construction timing, service level and quality, and pricing (pp. 27-31):

> The overall experience in Massachusetts with respect to each of these dimensions has been quite favorable...In almost all cases, cable systems have been constructed within the promised time period...problems with the level and quality of service are infrequent, and not terribly serious in nature...The pricing behavior exhibited by Massachusetts cable companies--increasing rates once every 2 or 3 years by an amount which in most cases is consistent with general price trends--does not seem to be unreasonably opportunistic.

Prager also sent out questionnaires to policymakers in 398 cabled communities nationwide. She obtained 216 responses and found that (p. 68):

> nearly seven-eights of those responding to question 35A [Would you advise another community similar to yours in size and composition to award a franchise to this firm?], would recommend their cable operators to other communities. While extreme responses...occur somewhat infrequently, extremely good ratings far outnumber extremely bad ratings for all aspects of firm behavior.

Only four percent of the respondents believed that their operator had asked for an excessive number of rate increases; six percent called the size of the rate increases excessive; six percent rated service quality as poor; and nine percent characterized their company as being "unresponsive" to community needs.

Prager's results are complemented by the findings of a nationwide, random-sample telephone survey of 66 systems coming onstream in the early 1980s (Zupan, 1987a). At the time the telephone survey was conducted in the fall of 1984, all of the surveyed system managers reported having dutifully adhered to their franchise agreements since the initiation of their respective contracts. And only two of the managers, albeit the managers of two of the largest systems, acknowledged that they were in the process of requesting revisions of any substance in their contracts: one system was seeking to turn over responsibility for local origination programming to the community in exchange for a slight increase in the firm's annual franchise fee; another system was asking for a 15-percent rate increase and a deferral in certain franchise fee payments. Both managers reported that their requests were predicated on grounds of economic necessity and expected the requests to be granted given "good to excellent" working relations with local policymakers.[16]

A survey of the trade press also reveals only a few cases of reneging by operators. If one excludes conflicts over changes in federal or state law,[17] rate increase requests which are contested but fall short of the inflation rate,[18] and minor changes in the nonprice provisions of a franchise contract,[19] an exhaustive examination of the trade press unearths fewer than 60 cases of operator reneging among the systems franchised between 1980 and 1986.[20] Relative to the total number of over 3000 systems franchised during that time period, the number 60 is quite small.[21]

Admittedly, evidence gleaned from the trade press that reneging by operators appears to be infrequent is not air-tight. For one thing, the observed reneges account for a disproportionate share of total national subscribership--since most of the cases involve large cities. It is also possible that problems in smaller cities receive less press or that many problems are settled internally by the parties involved and never publicly reported. Furthermore, there are obvious difficulties in classifying revisions as being either major or minor.

V.B. A Comparison of Predicted and Realized Penetration Rates

While the problems associated with a trade press survey cannot be glossed over, an alternative, more powerful, test for both the incidence of operator reneging and the reasons for such reneging is available; and the data necessary to perform the test are readily procurable. The alternative test compares the basic penetration rate (the ratio of basic subscribers to homes passed by cable in a community) predicted by the winning bidder at franchising time with the realized penetration rate.

Cities award cable franchises to the bidder promising the most desirable combination of construction timetable, prices, programming quality and politically salient nonprice concessions (cf., William Shew, 1984; and Prager, 1986). Nonprice concessions account for a sizable portion of cable costs and may include institutional networks linking the public facilities in a community, franchise fees and community programming.[22] Since nonprice concessions are primarily fixed costs, the expenditures a bidder pledges to make on them should depend on the expected profit level and thus, holding the promised price constant, on the bidder's projected rate of subscribership.

Given the criteria cities employ in awarding franchises, operator opportunism might well involve the following strategy:

Opportunistic Strategy #1

1. Begin by intentionally overestimating the penetration rate and the size of the expected profit "pie". Based on the inflated profit pie, promise a sufficient amount of nonprice concessions to win the franchise, even though you don't plan to deliver on all that you promise.
2. Once you win the contract and get your foot in the door, claim that you honestly overestimated the penetration rate and that the level of profits has turned out to be smaller than expected. Ask for a rate increase and scalebacks in the initial commitments you made to provide nonprice concessions.

The plausibility of Strategy #1 is attested to by the justifications operators typically proffer when requesting contract revisions. By far the most common rationale tendered is lower-than-anticipated subscribership. In Minneapolis, for instance, Rogers requested revisions because: "in late 1983 the company projected that it would have 70,401 [basic service] subscribers and 126,721 pay [e.g., Home Box Office, Showtime] units by now. As of July 31, 1985, Rogers had 34,348 subscribers and 39,300 pay units".[23] In Denver, Mile Hi's appeal was predicated on a basic penetration rate of 27 percent: "the company expected to have closer to 50 percent."[24]

An alternative opportunistic strategy for operators might entail the following:

Opportunistic Strategy #2

1. For the price initially promised the city, do not overestimate the penetration rate but intentionally underestimate costs.
2. Once you win the contract, exploit the city's investment in the relationship to obtain a rate increase and scalebacks in the initially promised level of nonprice concessions. Point to higher than projected costs to justify the contractual changes.

Finally, operator opportunism may involve the following tactic:

Opportunistic Strategy #3

1. For the price initially promised the city, do not overestimate the penetration rate or underestimate costs.
2. Once you win the contract, exploit the city's investment in the relationship to ask for a rate increase.

If the constraints on operator opportunism are weak and if any one of the three preceding strategies can be successfully employed by cable firms, two predictions can be made. First, anticipated market penetration should exceed realized penetration (with Strategies #2 and #3, this is because operators jack up prices after getting a foot in the door). Second, any variance in the amount by which expected penetration exceeds realized penetration should be due to factors affecting the rewards to opportunistic behavior. Other things being equal, that is, market overestimation should be greater where the gains from operator opportunism are larger.

These predictions are testable. The top panel of Table 1 compares expected and realized basic penetration rates for 1986 for all 68 systems franchised over 1979-1982 with: projected penetration rates on file at either the Massachusetts CATV Commission or the Cable Television Information Center (CTIC); and 1986 penetration levels published in trade periodicals.[25] The bottom panel makes a similar comparison for the 66 systems in telephone survey sample.[26]

As shown in Table 1, expected rates exceed their realized counterparts. The rates, however, are not dissimilar from one another in terms of statistical significance. Neither are they consistently dissimilar from one another when the sample is stratified by system size. As shown in Table 2, expected rates exceed realized rates only for large systems; by 6.3 percent for the 15 Massachusetts/CTIC systems passing over 50,000 homes and by 11.2 percent for the seven telephone survey systems with over 50,000 homes passed.[27] The operators of small systems predicted penetration virtually on the nose and, if anything, were conservative in their projections. For both samples, the operators of small systems "high-balled" their bids and underestimated subscribership!

The discrepancy between the small- and large-system experiences begs the question "Why?". One possible answer is that the gains from opportunism may be a positive function of system size. There may be, in other words, economies of scale to lying.[28] If large cities have weaker contract enforcement mechanisms or more substantial nonsalvageable costs (per home

TABLE 1

A COMPARISON OF EXPECTED AND REALIZED BASIC PENETRATION RATES

SAMPLE	EXPECTED PENETRATION RATE	REALIZED PENETRATION RATE	DIFFERENCE (t-stat.)	NO. OF OVER-PREDICTIONS (Sign stat.)
MASSACHUSETTS/ CTIC	50.1%	46.2%	3.9% (0.55)	32 out of 68 (-0.49)
TELEPHONE SURVEY	53.0%	46.3%	6.7% (0.67)	25 out of 53 (-0.41)

* Significant at the .05 level.
** Significant at the .01 level.

SOURCES: See text and Appendices A and B.
NOTE: Statistics concerning the number of overpredictions exclude cases of ties.

TABLE 2

A REEXAMINATION OF EXPECTED AND REALIZED BASIC PENETRATION RATES

SAMPLE	EXPECTED PENETRATION RATE	REALIZED PENETRATION RATE	DIFFERENCE (t-stat.)	NO. OF OVER-PREDICTIONS (Sign stat.)
MASSACHUSETTS/CTIC				
Large Cities	49.2%	42.9%	6.3% (1.26)	12 out of 15 (2.32)*
Small Cities	51.9%	53.0%	-1.1% (-0.33)	20 out of 53 (-1.79)
TELEPHONE SURVEY				
Large Cities	52.0%	40.8%	11.2% (1.45)	6 out of 6 (2.45)*
Small Cities	54.7%	55.5%	-0.8% (-0.10)	19 out of 47 (-1.31)

* Significant at the .05 level.
** Significant at the .01 level.

SOURCES: See text and Appendices A and B.
NOTE: Statistics concerning the number of overpredictions exclude cases of ties.

passed), it may be more rewarding for an operator to deceive a single city with 100,000 homes than to sucker 5 cities with 20,000 homes each. The same holds true if there are greater informational asymmetries or larger amounts of potentially-capturable consumer surplus (per home passed) in big cities.

On the other hand, the overestimation may reflect the lack of prior information available to bidders for big-city franchises. Due to the nature of federal regulations in the late 1960s and early 1970s, cable's growth in urban markets was stunted until the late 1970s (Bruce Owen, 1981; and Stanley Besen and Robert Crandall, 1981). In 1966, the Federal Communications Commission instituted a freeze on cable growth in the top 100 television markets. For the sake of protecting local broadcasting, cable operators in the top 100 markets were: ordered to carry all local stations; not allowed to import distant signals which duplicated local programming; and required to demonstrate that any proposed distant signal would not threaten even marginal local UHF stations. The FCC requirements left major urban markets dead in the water as far as cable was concerned (Besen and Crandall, 1981).

Due to the continued growth of the cable industry and the accompanying increase in the industry's political clout, the FCC began to thaw the freeze on cable activity in urban markets in 1972 (cf., Owen, 1981). As corroborated by Table 3, the regulatory senescence was paralleled by a boom in urban wiring. After 1974, the number and subscriber-share of large systems grew dramatically. The wiring of large cities was particularly concentrated, in a temporal sense, over the five-year period 1977-1982. As shown by Table 4, of the 152 large systems (systems passing over 50,000 homes) coming onstream between 1950 and 1982, almost 40 percent did so over 1977-1982.

When urban areas went up for bids, in clumped fashion, in the late 1970s and early 1980s, the bidders for these franchises had little prior evidence to base their market projections on; particularly evidence on the effect that

TABLE 3

BREAKDOWNS OF THE CABLE INDUSTRY BY SYSTEM SIZE FOR 1974 AND 1986

SYSTEM SIZE (BSUBS)	NO. OF SYSTEMS IN 1974	PERCENT BREAK-DOWN	NO. OF SYSTEMS IN 1986	PERCENT BREAK-DOWN
100000+	0	0.000	27	0.004
50000-99999	5	0.002	98	0.014
20000-49999	36	0.011	308	0.044
10000-19999	142	0.045	458	0.065
5000-9999	289	0.092	628	0.089
3500-4999	214	0.068	394	0.056
1000-3499	1035	0.329	1699	0.241
500-999	577	0.183	1106	0.157
0-499	851	0.270	2333	0.331
TOTAL NO. OF SYSTEMS	3149		7051	

SOURCE: Television and Cable Factbook (1974 and 1986).
NOTE: BSUBS is the number of basic subscribers.

TABLE 4

THE TIME PERIODS DURING WHICH BIG-CITY SYSTEMS CAME ONSTREAM

TIME PERIOD	NO. NEW FRANCHISES	NEW HP	NEW BSUBS	AVG. BPEN
1980-82	25	3648962	1403999	0.38
1977-79	32	4559999	2063110	0.45
1974-76	14	1476086	777289	0.53
1971-73	20	1999402	1121847	0.56
1968-70	20	2687459	1429348	0.53
1965-67	23	2273389	1344516	0.59
1962-64	11	1526333	844212	0.55
1959-61	1	73000	61778	0.85
1956-58	0	0	0	XX
1953-55	3	205283	116239	0.57
1950-52	3	465200	250000	0.54
TOTAL	152	18915113	9412338	0.50
1950-76	95	10706152	5945229	0.56
1977-82	57	8208961	3467109	0.42

SOURCE: Appendix C.

NOTES: Large cities are cities with over 50,000 homes passed by cable. Data for homes passed (HP), the number of basic subscribers (BSUBS), and basic penetration (BPEN) are for 1986.

substitutes for cable in urban areas, such as broadcast television (cf., Paul MacAvoy, 1977) and satellite master antenna television (miniature CATV systems found in apartment buildings and hotels) would have on subscribership.[29] An overreliance by big-city bidders on the penetration evidence from existing smaller systems may have produced the observed market misjudgments.

VI. DO MARKET MISCALCULATIONS REPRESENT OPERATOR OPPORTUNISM?

To investigate whether bidder's penetration miscalculations reflect weak weak contractual safeguards and prevalent bad behavior by cable firms or unanticipated changes in market information and well-working constraints on operator opportunism, econometric estimation is employed. The dependent variable to be examined is OVEREST: the difference between the projected and realized number of basic subscribers relative to the projected number of basic subscribers for the systems in the telephone survey sample.[30]

VI.A. Hypothesis #1: Prevalent Operator Opportunism

If cities' contractual safeguards are weak, the level of operator opportunism will depend on the profitability associated with guileful conduct. Hypothesis #1 thus predicts that operator opportunism will be more extensive the poorer the substitutes for cable in a community. Where substitutes are poorer, there is a large amount of consumer surplus (per home passed) to be captured by an operator through bad behavior.

The principal substitutes for basic cable service are broadcast television (cf., MacAvoy, 1977; Shew, 1984; and G. Kent Webb, 1983) and the satellite master antenna television (SMATV) systems (essentially miniature cable systems) found in hotels and apartment buildings (cf., Prager, 1986).

To measure the strength of broadcast television substitutes, the variable REGULARTV is employed. REGULARTV is the log of the number of network stations received (plus unity) in a cable system's locality—with account being made for signal quality according to the formula in Webb (1983).[31]

Two variables are utilized to reflect the competitiveness of SMATV: MDU and HOTELS. MDU is the percentage of a city's housing units accounted for by multiple-dwelling-unit structures with over 50 units. HOTELS is the percentage of a city's service-sector income deriving from hotels, motels and trailering parks.[32] If Hypothesis #1 is correct, REGULARTV, MDU and HOTELS should all have a negative impact on OVEREST. This is because, all other things being equal, the stronger are the substitutes for cable, the less consumer surplus (per home passed) there is to be gained by an operator through guileful conduct.

Holding the gains from opportunism constant, Hypothesis #1 also predicts that operator opportunism is apt to increase the greater the number of bidders there are for a franchise. This is because the winning bidder has to promise away more (in terms of both pricing and nonprice features) when there are more competitors vying for the franchise (Zupan, 1987a). Since more must be promised to get a foot in the door and to obtain the opportunity of appropriating a relationship's quasi-rents, OVEREST should be an increasing function of the number of BIDDERS for a franchise if Hypothesis #1 is correct.

A final factor which should heighten an operator's incentive to behave badly is the extent of a city's nonsalvageable costs in the franchise relationship. As discussed in Section II, the greater a city's irreversible investments, the larger are the quasi-rents that may be appropriated by a badly-behaving cable firm. Provided that the irreversible investments cities make in franchise relationships parallel the nonsalvageable costs sunk by operators, Hypothesis #1 thus predicts that OVEREST will be positively related to the size of operators' predetermined, per-home-passed level of costs (IDIOSYNK).[33] Projected penetration rates will exceed realized rates by a greater extent where cities are relatively more "stuck" with their operators.

VI.B. Hypothesis #2: Efficacious Enforcement Mechanisms and Unanticipated Changes in Market Information

In a regression model of market misjudgments, the coefficient associated with the variable BIDDERS should be insignificantly different from zero if the contractual enforcement mechanisms available to cities are potent. Of the five enforcement mechanisms discussed in Section IV above, variables reflecting the potency of four of them can be constructed.[34] These four are: cross-sectional reputation, within-contract reputation, cities' monopsony power, and vertical integration.

If multiple system ownership is desirable and if cities pay heed to prior track records when awarding franchises, an operator's incentive to behave badly will diminish. Where companies intend to parlay promise-keeping in current franchises into future business from new customers, market overestimations should be lower, all else equal. Consequently, the OVEREST observed in 1984 for a system in the telephone survey should be a negative function of the system's parent-MSO's growth rate over 1984-1986, in terms of total national basic subscribership.[35]

If within-contract reputation and cities' monopsony power restrain guileful operator conduct, the variable IDIOSYNK should have exactly the opposite sign predicted by Hypothesis #1.[36] Besides reflecting an operator's potential gains from opportunism, the predetermined per-home-passed level of cable costs also measures an operator's stake in the relationship. Over the franchise relationship's duration, the value of this stake depends critically on the joint effort the city makes to the relationship (a joint effort which may become suboptimal in response to brazen behavior by the operator) and is subject to monopsonistic expropriation by the city. Thus, if Hypothesis #2 is valid and if internal reputation and cities' monopsony power serve to keep operators honest, IDIOSYNK should have a negative effect on OVEREST.

By more closely aligning the interests of the parties to a contract, vertical integration may improve operator conduct. In the telephone survey sample, two systems are, in a broad sense, vertically integrated: one was built by the city's Utilities Board; the other is owned and operated by a local university. A dummy variable set equal to unity for these two systems (NONPROF) should have a negative effect on OVEREST if vertical integration circumscribes opportunistic conduct by operators.

Finally, Hypothesis #2 asserts that penetration miscalculations largely reflect changes in market information unforeseen by operators at franchising time. Among the unanticipated factors, the one perhaps most frequently pointed to in problem markets is the unexpected strength of substitutes for cable. In asking for concessions in Denver, for example, Mile Hi's president stated that "Although the cable television industry as a whole continues to prosper, we have discovered like other metropolitan markets, the consumer dollar in Denver is in competition with many entertainment options."[37] In Dallas, Warner faced significant problems in marketing multiple dwelling units due to competition from SMATV.[38] In Houston, the principal reason cited by Warner for a difficult operating environment, was the plethora of local broadcast signals.[39]

If operators' market misjudgments reflect the unforeseen strength of substitutes for cable in urban areas, the signs of the coefficients associated with the three variables REGULARTV, HOTELS, and MDU should be contrary to the signs predicted by Hypothesis #1. Provided that Hypothesis #2 is correct, REGULARTV, HOTELS, and MDU should have a positive impact on OVEREST.

VI.C. Estimation Procedure and Results

The equation of interest is the following:

[1]
$$\begin{aligned}\text{OVEREST} = {} & b0 + b1*\text{BIDDERS} + b2*\text{IDIOSYNK} + b3*\text{NONPROF} + b4*\text{HOTELS} + b5*\text{MDU} \\ & + b6*\text{REGULARTV} + b7*\text{REPUTATION} + b8*\text{YRSOP} + b9*\text{OUTLIER} + \text{ERROR}.\end{aligned}$$

The variable YRSOP is the log of the number of years (plus unity) that a system had been operating at the time its manager was interviewed. It is included in the regression to account for the fact that cable system penetration tends to approach projected levels at a diminishing rate (Webb, 1983) and that market misjudgments should therefore be smaller the longer a system has been in operation.[40] The variable OUTLIER is included because the smallest system in the sample, for which there was only one bidder, displayed the second highest OVEREST but was owned by the largest and fastest-growing MSO in the industry. Due to a substantial amount of prior information attesting to the MSO's reputation for building "plain vanilla" systems and building them well, it was deemed appropriate to adjust for this outlier.[41]

Tables 5 and 6 provide summary definitions and descriptive statistics for the Equation [1] variables. Of central interest in the explanation of operators' market miscalculations is the differential roles predicted by the two competing hypotheses for certain of the explanatory variables. In short:

TABLE 5

VARIABLE DEFINITIONS

OVEREST	Expected minus realized number of basic subscribers divided by the expected number of basic subscribers.
BIDDERS	The number of companies submitting bids for a city's franchise.
IDIOSYNK	The predetermined monthly level of non-franchising-related costs (capital and operating) per home passed by a system. See Zupan (1987a) for further details.
NONPROF	A dummy variable set equal to unity if a system is nonprofit (2 systems).
HOTELS	The percentage of a city's service-sector income accounted for by hotels, motels and trailering parks.
MDU	The percentage of a city's housing units that are in multiple-dwelling-unit structures having at least 50 units.
REGULARTV	Log of the number of network stations received (plus unity) in the cable system's locality--with account being made for the quality of the received signal.
REPUTATION	The growth over 1984-1986, in terms of the number of basic subscribers, of the parent-MSO of the winning bidder.
YRSOP	Log of the number of years (plus unity) that a system has been operating.
HOMES	The number of homes in a community.
OUTLIER	A dummy variable set equal to unity for the smallest system in the sample which displayed the second highest OVEREST but was owned by the largest and fastest-growing MSO in the industry--a MSO with a reputation for building "plain vanilla" systems and building them well.

TABLE 6

DESCRIPTIVE STATISTICS FOR THE VARIABLES EMPLOYED IN THE ECONOMETRIC ANALYSIS

VARIABLES	MEAN	STD. DEV.
OVEREST	-0.015	0.254
BIDDERS	4.439	3.774
IDIOSYNK	116.999	22.773
NONPROF	0.030	0.173
HOTELS	0.081	0.128
MDU	0.032	0.043
REGULARTV	1.058	0.343
REPUTATION	0.012	0.022
YRSOP	1.150	0.262
HOMES	27475.300	55062.900
OUTLIER	0.015	0.015

PREDICTED EFFECTS

VARIABLES	HYPOTHESIS #1: Prevalent Operator Opportunism	HYPOTHESIS #2: Potent Contractual Safeguards and Unanticipated Information
BIDDERS	+	0
IDIOSYNK	+	-
NONPROF	0	-
HOTELS	-	+
MDU	-	+
REGULARTV	-	+
REPUTATION	0	-

The results of an ordinary least-squares estimation of Equation [1] are reported in Column 1 of Table 7. As can be seen, the results provide no support for Hypothesis #1. The size of OVEREST does not vary positively with the number of bidders. In addition, operators underestimated market size where there were more quasi-rents (IDIOSYNK) to be appropriated. Finally, bidders underestimated market penetration where a community had poorer substitutes for cable and where there was a greater amount, per home passed, of consumer surplus that might have been appropriated through opportunistic behavior.[42]

The Column 1 results are thoroughly consistent with Hypothesis #2. Contractual safeguards such as cross-sectional and internal reputation and cities' monopsony power appear to be especially effective at limiting bad behavior by operators. Overestimates of market size, furthermore, appear to reflect the unforeseen factors cited by operators: namely, stronger-than-expected competition from broadcast television and SMATV.

VI.D. Accounting for Possible Economies of Scale to Lying

Despite the strong results of Column 1, there remains the nagging possibility that for some as-of-yet unspecified reason there are economies of

TABLE 7
AN ECONOMETRIC EXAMINATION OF OPERATORS' MARKET MISCALCULATIONS

EXPLANATORY VARIABLES	COLUMN 1	COLUMN 2
CONSTANT	0.102 (0.54)	0.141 (0.75)
BIDDERS	-0.004 (-0.54)	-0.008 (-0.94)
IDIOSYNK	-0.003** (-2.75)	-0.003** (-2.68)
NONPROF	-0.678** (-5.12)	-0.667** (-5.09)
HOTELS	1.119** (5.35)	1.091** (5.26)
MDU	0.908 (1.60)	0.762 (1.33)
REGULARTV	0.484** (5.60)	0.454** (5.16)
REPUTATION	-0.273* (-2.11)	-0.226* (-1.71)
YRSOP	-0.321** (-3.61)	-0.339** (-3.81)
OUTLIER	1.038** (4.21)	0.971** (3.90)
HOMES	--	0.665 (1.47)
RSQ	0.59	0.60
CRRSQ	0.52	0.53

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: Dependent variable in the ordinary least-squares regressions is OVEREST. T-statistics are in parentheses.

scale to deceit. For example, the effects of REGULARTV, HOTELS and MDU on OVEREST may be due to these explanatory variables "standing in" for system size and that it is more profitable for an operator to deceive a large city with 100,000 homes than to sucker 5 cities with 20,000 homes each.

To investigate this possibility, the number of homes in a community is added as an explanatory variable to the Equation [1] model. If there are economies of scale to deceit, HOMES should have a positive effect on OVEREST and the addition of HOMES should drive the results towards the outcome predicted by Hypothesis #1.

As Column 2 of Table 7 shows, the results hardly budge when HOMES is added to the explanation of market miscalculations. The estimated coefficient on HOMES is positive but insignificant. Overestimations appear to have occurred specifically where there are strong substitutes for cable, as operators of problem systems have claimed, and not more generally as a function of city size.

VI.E. Supplementary Evidence

The econometric findings that contractual safeguards are effective and that operators' market misjudgments are not predicated on dishonesty would be most strongly corroborated by proof that in franchise competitions conducted after 1982 operators learned from past mistakes. Unfortunately, relatively few large cities have been franchised since 1982 and data on how accurately the winning bidders for these cities predicted market size are largely still unavailable.[43] The available anecdotal evidence, however, is consistent with the foregoing econometric results.

To begin with, the reported contract revisions in cities franchised after 1982 have occurred almost exclusively before the start of system construction.

For example, before wiring a single mile of Washington D.C., franchisee DCI requested a stop on all franchise fee payments, a channel downgrade, a basic rate hike of $2, and millions of dollars less in public access funding. DCI termed the existing contract "economically unviable", downgraded projected basic penetration from 42 to 30 percent, and rationalized the requested revisions by citing the same problems mandating franchise revisions in other large cities.[44] Before beginning operations in Detroit, Barden obtained concessions of more than $25 million from an initial $178 million commitment.[45] To justify the scalebacks, Barden claimed "new information on the economics of big-city cable systems."[46]

Revisions requested antecedent to system construction are apt to reflect the arrival of unanticipated information. Because the costs associated with the franchise award process are small relative to the total investment parties make in a typical cable franchise relationship, attempted bad behavior by operators immediately subsequent to the award is unlikely to be successful.[47] An aggrieved city can terminate the contract at comparatively little cost and transfer the award to the next highest bidder. Cities' acquiescence to most of the revisions requested prior to system construction may thus reflect the arrival of new, less optimistic information on the viability of cable in urban markets--information equally discouraging to all potential operators.

Another possible sign that operators may be learning from past errors is the recent decline in the number of bidders for urban franchises. The 1984 competition for Baltimore, for example, drew only one bidder. Four other MSOs looked at the city and decided not to bid. A spokesman for one of the MSOs termed the franchise economically unviable and stated that "We have no intention of proposing to build something in the bid and then changing it after we win the franchise."[48] Sections of Chicago, Los Angeles, New York, and Philadelphia similarly have drawn only one bidder. On average, there were 2.9 bidders for the 27 large cities franchised after 1982 for which data are available. In contrast, the number of bidders for the 20 large telephone survey or Massachusetts/CTIC systems franchised over 1979-82 averaged 6.7.[49]

Finally, provisions intended to promote the spirit of long-term, incomplete cable franchise contracts in the face of market uncertainties have begun to surface. For example, DCI's renegotiated Washington D.C. contract makes the size of community programming payments contingent on the basic penetration rate.[50] Cablevision's revised Boston contract defers franchise fee payments until basic penetration reaches 50 percent.[51] The renewal contract negotiated in 1984 in Oakland specifies that an institutional network need not be built until the operator achieves a 15 percent return on investment.[52]

VII. SUMMARY AND CONCLUSION: THEY 'AINT MISBEHAVIN'

In analyzing the Oakland cable franchising case, Williamson (1976) emphasizes two basic themes. First, because of their incompleteness, long-term cable franchise contracts resemble direct regulation in practice. Second, in cases more complex than Harold Demsetz's (1968) automobile license plate example, problems such as insufficient competition at bidding time, incumbency advantages at refranchising time, and operator opportunism during contract execution render franchise bidding schemes suspect as mechanisms for promoting efficiency.

The first of these assertions is easily accepted. Cable franchise contracts are not fully contingent, once-and-for-all accords. The winning bidder does not disappear after winning the franchise and resurface 15 years later at refranchising time. Rather, there are many details to be ironed out along the way and numerous unanticipated events to be mutually dealt with. And, as Klein and Kenney (1986) have argued, it may be excessively costly to attempt to iron out the details beforehand and to specify all parties' contractual obligations under all possible states of the world that may attain. It may be desirable for franchise contracts to resemble direct regulation in practice.

Williamson's second assertion is more debatable and is the subject of systematic scrutiny by this and two companion pieces (Zupan, 1987a and 1987b). The results of the scrutiny suggest that even in more complex cases such as CATV, franchise bidding schemes may work far better than is commonly believed.

One of the companion pieces (Zupan, 1987a) investigates the extent to which franchise bidding schemes are successful at controlling monopoly pricing. The results indicate that the restraint is substantial and that costs and prices would be even further minimized if cities did not pursue certain nonprice concessions when conducting franchise competitions.

This paper and the second companion piece (Zupan, 1987b) focus on the problem of opportunism and find that cable operators behave much better than is commonly believed. Reneging on promises is infrequent and when it occurs it appears largely to reflect unanticipated changes in market information. The advantages of incumbency, furthermore, are typically not "milked" by operators at the time of franchise renewal (Zupan, 1987b).

The institutional mechanisms which appear to so effectively restrain guileful operator conduct are cross-sectional and within-contract reputational considerations and cities' monopsony power. Were it not for concerns over having a "bad name" and the incumbency advantages available to cities, operators would find it opportune to behave opportunistically.

FOOTNOTES

[1] Multichannel News, January 20, 1986, p. 24, and March 10, 1986, p. 31.

[2] CableVision, June 10, 1985, p. 29, and August 12, 1985, p. 22.

[3] Multichannel News, May 20, 1985, p. 15.

[4] CableVision, June 10, 1985, p. 29, and Multichannel News, February 18, 1985, p. 6.

[5] Multichannel News, December 10, 1984, p. 39.

[6] See, for example, CableReports, December 1980, p. 1.

[7] These arguments should not be taken to imply that contractual enforcement is, in general, undesirable. When a contract involves little joint effort and where the parties are aware of the risks involved and can easily define their individual duties under particular outcomes, *ex post* revisions should be unwarranted *per se*. Matters differ, however, when considerable joint effort is required and where the risks and duties are impossibly complex and/or costly to specify; adherence to the spirit of the initial contract may be far more appropriate than *per se* enforcement of the initial provisions (Klein and Kenney, 1986; and Posner, 1986).

[8] For empirical evidence on the importance that considerations of fairness play in the face of unforeseen market developments, see Daniel Kahneman, Jack Knetsch, and Richard Thaler (1986).

[9] The data come from Cable TV Franchising, April 14, 1983, p. 6. While the infrequency of municipal ownership may indicate that opportunism is not a severe problem, it may also be due to factors such as: learning-by-doing advantages possessed by private firms; the monopsony power multiple system operators are able to exercise in input markets; and/or ideological considerations (cf., Jeffrey Dubin and Peter Navarro, 1987).

[10] The top 100 MSOs alone accounted for 91 percent of the total subscribership (CableFile, 1986, p. I-5).

[11] Even after all communities are franchised, cross-sectional reputation proves valuable provided that there are a nontrivial number of system sales and that cities look at firms' performance records when deciding whether or not to approve a sale. In 1986, system sales totaled $8 billion and 15 percent of all subscribers changed company hands (CableVision, February 16, 1987, p. 54).

[12] By state law, all Massachusetts communities are required to file a report specifying the reasons for approving or denying a bid. These reports are on file with the Massachusetts CATV Commission.

[13] The investments parties make during a 15-year franchise do not represent an infinite-horizon game and may be subject to the chain-store paradox (Reinhard Selten, 1975). Scherer's argument may nonetheless be valid if there is uncertainty over when one's partner will make investments (Paul Milgrom and John Roberts, 1983, and Drew Fudenberg and Jean Tirole, 1986) or actions taken in the game's final period have cross-sectional reputation value.

[14] CableAge, April 2, 1984, p. 19.

[15] For example, a city's refusal to allow cable prices to keep up with inflation might be classified as opportunism on the part of the city.

[16] Both requests were approved without contention by the cities.

[17] For example, with the recently-passed federal law which prohibits local rate control effective January 1, 1987, legal disputes have arisen over whether existing contracts are grandfathered.

[18] As documented by Cable TV Franchising, most rate increase requests are fully approved without contention by cities.

[19] Examples of minor disputed contract changes are attempts by operators to drop the carriage of a single, local broadcast signal or to add a controversial pay service such as the Playboy Channel.

[20] The trade periodicals examined were Cable TV Franchising, CableVision, Multichannel News and CableReports.

[21] Cable Update, December 31, 1979, p. 1, reveals that there were 4048 systems as of December 31, 1979. By 1986, the number of systems had grown to 7051 (CableFile, 1986).

[22] According to Zupan (1987a), politically salient nonprice concessions account for 26 percent of construction costs and 11 percent of operating costs.

[23] Multichannel News, September 2, 1985, p. 7.

[24] Cablevision, September 23, 1985, p. 22.

[25] The projections were made at franchising time. The CTIC files include proposal data from all the cities to which CTIC has provided consulting assistance. Realized penetration rates are from CableFile (1986) and TV and Cable Factbook (1986). Reported percentages are weighted averages based on the Appendix A data—with the weights depending on the number of homes passed by each system. Basic penetration receives primary focus because pay penetration data are less consistently available. Where available, however, the pattern of the difference between expected and realized pay penetration rates mirrors the reported pattern for basic penetration rates.

[26] Projected rates are for October 1984 and were provided by the surveyed operators. Reported percentages are weighted averages based on the Appendix B data.

[27] Systems passing over 50,000 homes are "large" since the average system in the U.S. in 1986 passed 8,265 homes (CableFile, 1986).

[28] This hypothesis does not receive support from the available data on how much winning bidders spend to obtain a franchise (Zupan, 1987a). Other things being equal (such as the number of bidders), a winning bidder's application expenditures are a linear function of the number of homes in the community passed by cable.

[29] In vying for large cities over 1977-1982, bidders did have information on subcribership in urban systems franchised prior to 1977. As shown in Appendix D, this information perhaps should have signalled trouble. For the 36 large cities franchised prior to 1974, the average penetration rate was 37 percent in 1976--much lower than the concurrent national average of 53 percent. The Oakland system analyzed by Williamson is one of the large cities in this sample and has the second lowest penetration rate--15 percent. While bidders over 1977-1982 did have information on existing subscribership rates, they had no way of knowing whether the lower urban rates were attributable to federal regulations or to other factors such as healthier substitutes for cable.

[30] Data for key variables in the analysis below are not available for systems in the Massachusetts/CTIC sample. Consequently, the econometric analysis focuses on the telephone survey sample of systems.

[31] REGULARTV is in log form to account for the diminishing value associated with each additional broadcast signal (Webb, 1983). Signal quality is accounted for by noting a system's proximity to a broadcast signal's transmitter relative to the system's distance to the signal's Grade B contour.

[32] MDU and HOTELS are based on Census data.

[33] IDIOSYNK reflects both capital and operating costs, does not include franchising-related costs, and is constructed according to the method employed by Zupan (1987a).

[34] Data reflecting the severity of penalty clauses are unavailable.

[35] There is a potential simultaneity problem since REPUTATION influences OVEREST, but a market misjudgment made in 1984 may affect the parent-MSO's 1984-1986 growth rate. Since all of the telephone survey systems had kept their franchise promises, however, the probability of a simultaneity problem is not high. Given the consistently good conduct, the misjudgments made by the surveyed systems are unlikely to have influenced their parent-MSOs' 1984-1986 growth rates--growth rates are determined primarily by new franchise awards and prior track records play a key role in city's award decisions. REPUTATION is preferred to an alternative measure reflecting the total subscriber size, in 1984, of a system's parent-MSO. This is since the alternative measure may represent either a desire for future growth or brand-name capital that may be milked in the current period. Appendix E reports the results obtained by employing REPUTATIONALT. As expected, REPUTATIONALT has no effect on OVEREST.

[36] As an additional measure of the potency of internal reputation and cities' monopsony power, the length of a system's franchise contract proves to be insignificant when it is employed as an explanatory variable. This may be due to the limited amount of variation in franchise length across systems--58 of the 66 telephone survey systems had durations of 15 years.

[37] Multichannel News, December 23, 1985, p. 6.

[38] Cablevision, November 22, 1982, p. 228.

[39] Cablevision, June 23, 1986, p. 60.

[40] YRSOP is also included in the regression to account for the possibility that managers may have provided benchmark, fifth-year projected penetration levels instead of the penetration rates expected for the fall of 1984.

[41] The largest MSO owned only one system in the sample. For representative information on this MSO's reputation, see "The Plain Vanilla Juggernaut" (Cablevision, March 18, 1985, pp. 28-33). The results obtained when the outlier is unaccounted for are reported in Appendix F. All results of interest are unchanged except for the sign and significance of the REPUTATION variable.

[42] Appendix G reports the results of an analogous examination of pay penetration mistakes. The results are, if anything, supportive of Hypothesis #2 although there appears to be greater noise in the POVEREST variable. The lower explanatory power may be due to missing variables (data on VCR competitiveness across cities, for example, are unavailable). Operators also admitted that pay penetration is harder to predict than basic penetration and that less stress was placed on expected pay penetration when bidding for a franchise--perhaps because pay services account for a smaller share of total revenues (26 percent in 1982).

[43] After 1982, only a few major cities remained unwired. These cities included Cleveland, Baltimore, Detroit, Philadelphia, Chicago, and portions of New York and Los Angeles. Realized penetration rates for almost all of the recently awarded big-city franchises are as of yet unlisted in CableFile or the Television and Cable Factbook--primarily because construction has been delayed due to contract renegotiations (see text below).

[44] Multichannel News, June 24, 1985, p. 6.

[45] Multichannel News, December 23, 1985, p. 2.

[46] Multichannel News, April 1, 1985, p. 12.

[47] See, for example, Zupan (1987a).

[48] Cablevision, November 15, 1982, p. 20.

[49] The difference is statistically significant. The t-statistic associated with the hypothesis that the two sample means (from independent samples) differ is 5.98.

[50] Multichannel News, December 3, 1984, p. 8.

[51] Multichannel News, November 11, 1985, p. 7.

[52] Cable TV Franchising, March 6, 1984, p. 8.

APPENDIX A

EXPECTED AND REALIZED BASIC PENETRATION RATES FOR THE MASSACHUSETTS/CTIC SAMPLE

CITY	STATE	HOMES PASSED	EXPECTED BPEN	REALIZED BPEN
Dallas	TX	323848	0.43	0.28
Cincinnati	OH	302240	0.51	0.49
Kansas City	MO	290000	0.45	0.38
Boston	MA	250000	0.41	0.28
Omaha	NE	125000	0.60	0.55
Jefferson County	KY	116620	0.57	0.65
Tucson	AZ	115000	0.49	0.45
Covington	KY	89225	0.53	0.48
Lexington	KY	86114	0.59	0.63
Garland	TX	83058	0.50	0.48
Arlington	TX	71000	0.50	0.42
Fort Wayne	IN	62500	0.61	0.47
Eden Prarie	MN	60878	0.61	0.40
Vancouver	WA	57500	0.49	0.44
Springfield	MA	54000	0.42	0.67
Scottsdale	AZ	46704	0.45	0.43
Roseville	MN	44850	0.54	0.24
Wichita Falls	TX	43100	0.71	0.52
Irving	TX	39000	0.69	0.44
Fall River	MA	38826	0.44	0.51
Brockton	MA	34000	0.50	0.46
Glendale	AZ	32649	0.50	0.49
Green Bay	WI	32500	0.45	0.55
Farmington	MI	28345	0.55	0.63
Lynn	MA	27500	0.34	0.48
Fargo	ND	26157	0.51	0.52
Newton	MA	26000	0.45	0.70
St. Louis County	MO	25000	0.50	0.42
Brookline	MA	24729	0.41	0.32
Dracut	MA	22636	0.48	0.56
Richardson	TX	22000	0.30	0.46
Sheboygan	WI	20900	0.50	0.71
Chicopee	MA	20500	0.72	0.78
Bowling Green	KY	20177	0.60	0.80
Weymouth	MA	19500	0.55	0.53
Denton	TX	19010	0.54	0.56
Peabody	MA	18500	0.45	0.65
Arlington	MA	18400	0.48	0.55
Longmont	CO	17000	0.52	0.49

APPENDIX A
EXPECTED AND REALIZED BASIC PENETRATION RATES FOR THE MASSACHUSETTS/CTIC SAMPLE

CITY	STATE	HOMES PASSED	EXPECTED BPEN	REALIZED BPEN
Beverly	MA	16491	0.50	0.63
Taunton	MA	16026	0.46	0.58
Gloucester	MA	15500	0.48	0.59
Plymouth	MA	15000	0.35	0.51
Thornton	CO	14900	0.53	0.49
Chapel Hill	NC	14117	0.59	0.65
Littleton	CO	14100	0.59	0.43
Newburyport	MA	13465	0.52	0.65
Manitowoc	WI	13115	0.53	0.64
Watertown	MA	13020	0.50	0.45
Attleboro	MA	12503	0.52	0.60
Northampton	MA	12000	0.60	0.63
Braintree	MA	12000	0.62	0.59
Lewisville	TX	11980	0.64	0.51
Inkster	MI	11544	0.54	0.40
Garden City	MI	11300	0.51	0.64
Natick	MA	11290	0.43	0.52
Brunswick	OH	10863	0.46	0.52
Marlborough	MA	10293	0.73	0.66
Fairhaven	MA	9933	0.56	0.66
Lexington	MA	9700	0.43	0.65
Milton	MA	8600	0.53	0.55
Wakefield	MA	8200	0.40	0.63
Germantown	TN	8171	0.58	0.60
Columbia Heights	MN	7800	0.65	0.38
Hastings	MN	4700	0.63	0.51
Steamboat Springs	CO	4000	0.71	0.88
Chaska	MN	3500	0.60	0.35
Dodgeville	WI	2343	0.64	0.52
TOTAL		3071420		
AVERAGE		45168	0.50	0.46

SOURCES: Massachusetts Community Antenna Television Commission and the CTIC.
NOTE: BPEN stands for basic penetration.

APPENDIX B
EXPECTED AND REALIZED BASIC PENETRATION RATES FOR THE TELEPHONE SURVEY SAMPLE

SYSTEM NO.	HOMES PASSED	EXPECTED BPEN	ACTUAL BPEN
1	318172	0.50	0.35
2	197000	0.50	0.36
3	145000	0.50	0.47
4	120000	0.46	0.42
5	90000	0.62	0.44
6	62500	0.78	0.62
7	59000	0.43	0.43
8	45672	0.74	0.74
9	45142	0.56	0.56
10	34000	0.46	0.46
11	33200	0.52	0.58
12	26700	0.45	0.41
13	26000	0.50	0.48
14	23500	0.49	0.56
15	22636	0.63	0.56
16	19500	0.50	0.53
17	18500	0.48	0.50
18	18184	0.47	0.47
19	18000	0.55	0.64
20	15000	0.51	0.55
21	14100	0.55	0.62
22	13500	0.55	0.65
23	13255	0.48	0.59
24	12582	0.53	0.47
25	11800	0.50	0.51
26	11500	0.50	0.50
27	11000	0.58	0.49
28	10020	0.50	0.40
29	9900	0.52	0.53
30	9700	0.65	0.65
31	9050	0.72	0.72
32	9050	0.55	0.66
33	8600	0.50	0.55
34	8541	0.40	0.35
35	7626	0.65	0.42
36	6400	0.50	0.59
37	6360	0.70	0.32
38	6000	0.59	0.59
39	5500	0.50	0.68
40	4511	0.60	0.79

APPENDIX B

EXPECTED AND REALIZED BASIC PENETRATION RATES FOR THE TELEPHONE SURVEY SAMPLE

SYSTEM NO.	HOMES PASSED	EXPECTED BPEN	ACTUAL BPEN
41	4500	0.70	0.78
42	4400	0.57	0.67
43	3850	0.70	0.62
44	3185	0.65	0.35
45	2734	0.50	0.50
46	2725	0.55	0.55
47	2550	0.60	0.61
48	2200	0.60	0.45
49	1900	0.60	0.60
50	1722	0.45	0.84
51	1700	0.50	0.65
52	1637	0.80	0.81
53	1548	0.50	0.59
54	1500	0.33	0.64
55	1302	0.28	0.28
56	1200	0.70	0.60
57	1025	0.60	0.72
58	980	0.60	0.31
59	968	0.51	0.40
60	875	0.50	0.44
61	830	0.72	0.67
62	800	0.68	0.81
63	575	0.50	0.65
64	565	0.50	0.64
65	500	0.50	0.66
66	325	0.70	0.42
TOTAL	1572797		
AVERAGE	23830	0.53	0.46

SOURCES: See text and Zupan (1987a).
NOTE: BPEN stands for basic penetration.

APPENDIX C
AWARD DATES AND BASIC PENETRATION RATES FOR LARGE-CITY SYSTEMS

NO.	CITY	STATE	AWARD	BEGAN	HP	BSUBS	BPEN
1	Denver	CO	1982	1983	194200	54400	0.28
2	Minneapolis	MN	1982	1983	131815	39731	0.30
3	Tampa	FL	1982	1983	76000	29618	0.39
4	Torrance	CA	1982	1983	55000	25668	0.47
5	Boston	MA	1981	1982	250000	70876	0.28
6	Fort Worth	TX	1981	1982	140000	54910	0.39
7	Fresno	CA	1981	1981	137000	46164	0.34
8	Indianapolis-ATC	IN	1981	1981	141600	62200	0.44
9	Santa Ana	CA	1981	1982	65000	21136	0.33
10	Springfield	MA	1981	1982	54000	36000	0.67
11	Tucson	AZ	1981	1982	115000	51500	0.45
12	Vancouver	WA	1981	1982	57500	25500	0.44
13	Arlington	TX	1980	1981	71000	30000	0.42
14	Atlanta	GA	1980	1981	312000	123215	0.39
15	Cincinnati	OH	1980	1980	302240	147000	0.49
16	Dallas	TX	1980	1982	323848	90150	0.28
17	Little Rock	AR	1980	1980	73208	41601	0.57
18	Los Angeles-W Val	CA	1980	1981	167364	57000	0.34
19	New Orleans	LA	1980	1982	217193	76107	0.35
20	Newark	NJ	1980	1982	117000	20000	0.17
21	Oklahoma City	OK	1980	1980	171994	62247	0.36
22	Omaha	NE	1980	1981	125000	69000	0.55
23	Pittsburgh	PA	1980	1980	165000	74000	0.45
24	Portland-E	OR	1980	1982	136000	66438	0.49
25	Trenton	NJ	1980	1981	50000	29538	0.59
26	Anaheim	CA	1979	1980	64645	28556	0.44
27	Anchorage	AK	1979	1980	65000	33000	0.51
28	Clearwater	FL	1979	1980	154000	88000	0.57
29	Evanston/Oak Park	IL	1979	1979	135000	54000	0.40
30	Hollywood	FL	1979	1980	52870	28338	0.54
31	Houston-NE	TX	1979	1980	153255	43676	0.28
32	Houston-WAmex	TX	1979	1981	485000	191100	0.39
33	Huntington Beach	CA	1979	1980	95000	45590	0.48
34	Kansas City	MO	1979	1980	290000	110300	0.38
35	Las Vegas	NV	1979	1980	160000	62000	0.39
36	Lexington	KY	1979	1980	86114	53956	0.63
37	Los Angeles-CC	CA	1979	1980	295644	32364	0.11
38	Louisville	KY	1979	1980	142812	71117	0.50
39	Nashville	TN	1979	1980	179500	65632	0.37
40	Providence	RI	1979	1982	76000	33000	0.43
41	Richmond	VA	1979	1980	52029	33800	0.65
42	Springfield	MO	1979	1979	61245	30912	0.50
43	Boise	ID	1978	1979	80910	33200	0.41
44	Cedar Rapids	IO	1978	1979	51500	33000	0.64
45	Fort Lauderdale	FL	1978	1979	110000	54000	0.49

APPENDIX C
AWARD DATES AND BASIC PENETRATION RATES FOR LARGE-CITY SYSTEMS

NO.	CITY	STATE	AWARD	BEGAN	HP	BSUBS	BPEN
46	Fort Wayne	IN	1978	1979	62500	29648	0.47
47	Jacksonville	FL	1978	1979	232504	121475	0.52
48	Lowell	MA	1978	1979	54000	29470	0.55
49	North Miami	FL	1978	1978	137222	81449	0.59
50	Rochester	NY	1978	1980	239300	123500	0.52
51	San Antonio	TX	1978	1979	375903	200112	0.53
52	Syracuse	NY	1978	1979	66691	39133	0.59
53	Wichita	KA	1978	1979	131843	83080	0.63
54	Albuquerque	NM	1977	1978	144000	56770	0.39
55	Hartford	CN	1977	1977	113000	63173	0.56
56	Indianapolis-Wab	IN	1977	1979	152985	83621	0.55
57	Mesa	AZ	1977	1978	59527	26138	0.44
58	Bridgeport	CN	1976	1977	93000	46866	0.50
59	Chattanooga	TN	1976	1977	148200	58086	0.39
60	Dayton	OH	1976	1977	85900	38200	0.44
61	Evansville	IN	1976	1979	65600	41405	0.63
62	Memphis	TN	1976	1976	267000	115000	0.43
63	Montgomery	AL	1976	1977	61420	41827	0.68
64	Yonkers	NY	1976	1977	60000	26688	0.44
65	Roanoke	VA	1975	1976	54012	36248	0.67
66	Baton Rouge	LA	1974	1975	129222	75000	0.58
67	Lansing	MI	1974	1975	80720	49687	0.62
68	Norfolk	VA	1974	1978	172021	105000	0.61
69	Portland	ME	1974	1975	53819	38300	0.71
70	Shreveport	LA	1974	1976	101800	53800	0.53
71	Spokane	WA	1974	1976	103372	51182	0.50
72	Arlington	VA	1973	1978	78000	34000	0.44
73	Birmingham	AL	1973	1976	120000	54600	0.46
74	Charleston	SC	1973	1973	113286	56820	0.50
75	Columbus-C	OH	1973	1973	104834	57408	0.55
76	Des Moines	IA	1973	1974	118000	59564	0.50
77	Durham	NC	1973	1977	55300	29600	0.54
78	Grand Rapids	MI	1973	1976	150000	87153	0.58
79	Honolulu	HA	1973	1973	163300	98500	0.60
80	Knoxville	TN	1973	1975	106000	56500	0.53
81	Rockford	IL	1973	1973	86500	49000	0.57
82	Saint Petersburg	FL	1972	1972	115792	70000	0.60
83	Scranton	PA	1972	1973	74260	44790	0.60
84	Stockton	CA	1972	1973	90000	50000	0.56
85	W. Palm Peach	FL	1972	1972	62172	51832	0.83
86	Waterbury	CN	1972	1974	54847	40417	0.74
87	Columbus-ATC	OH	1971	1973	74600	34500	0.46
88	Corpus Christi	TX	1971	1972	62309	36155	0.58
89	Overland Park	KA	1971	1971	76281	46951	0.62
90	Saginaw	MI	1971	1973	50187	25879	0.52

APPENDIX C
AWARD DATES AND BASIC PENETRATION RATES FOR LARGE-CITY SYSTEMS

NO.	CITY	STATE	AWARD	BEGAN	HP	BSUBS	BPEN
91	Tulsa	OK	1971	1974	243734	138178	0.57
92	Amarillo	TX	1970	1970	63516	38500	0.61
93	Columbus-W	OH	1970	1971	141634	64500	0.46
94	Jackson	MS	1970	1972	95000	56700	0.60
95	Manhattan-GW	NY	1970	1971	258242	112000	0.43
96	Manhattan-MC	NY	1970	1971	415800	219600	0.53
97	Modesto	CA	1970	1970	52157	39198	0.75
98	Orlando	FL	1970	1971	242100	128100	0.53
99	Worcester	MA	1970	1970	71450	35076	0.49
100	Oakland	CA	1969	1970	115000	31000	0.27
101	Pompano Beach	FL	1969	1969	163169	83359	0.51
102	Savannah	GA	1969	1969	69800	38600	0.55
103	Winston-Salem	NC	1969	1970	88283	58035	0.66
104	Akron	OH	1968	1970	150000	60075	0.40
105	Albany	NY	1968	1974	85100	55300	0.65
106	Augusta	GA	1968	1970	70787	49854	0.70
107	El Paso	TX	1968	1972	123684	61000	0.49
108	Greensboro	NC	1968	1968	62300	38200	0.61
109	Los Angeles-GW	CA	1968	1969	213377	139051	0.65
110	Mobile	AL	1968	1969	70660	36400	0.52
111	Wilmington	DE	1968	1970	135400	84800	0.63
112	Charlotte	NC	1967	1967	131400	78700	0.60
113	Colorado Springs	CO	1967	1969	118200	51300	0.43
114	Lincoln	NE	1967	1968	78572	53169	0.68
115	Lubbock	TX	1967	1967	66000	30033	0.46
116	Pensacola	FL	1967	1969	69927	44061	0.63
117	Raleigh	NC	1967	1968	85800	48000	0.56
118	San Jose	CA	1967	1968	216000	110000	0.51
119	Bakersfield	CA	1966	1966	82207	47843	0.58
120	Buffalo	NY	1966	1966	140740	65000	0.46
121	Lancaster	PA	1966	1966	60348	42583	0.71
122	Newport News	VA	1966	1967	50000	35000	0.70
123	Peoria	IL	1966	1973	80000	48550	0.61
124	Rockford	IL	1966	1973	86500	49000	0.57
125	Salt Lake City	UT	1966	1970	62658	52968	0.85
126	Toledo	OH	1966	1966	150000	102572	0.68
127	Brookhaven	NY	1965	1966	59000	48500	0.82
128	Canton	OH	1965	1966	93637	53046	0.57
129	Cleveland Heights	OH	1965	1966	125600	69411	0.55
130	Flint	MI	1965	1966	126581	70900	0.56
131	Harrisburg	PA	1965	1965	73000	58130	0.80
132	Madison	WI	1965	1973	119000	62639	0.53
133	San Bernardino	CA	1965	1968	51019	30911	0.61
134	San Diego-SW	CA	1965	1965	147200	92200	0.63
135	Binghampton	NY	1964	1964	73472	56147	0.76

APPENDIX C
AWARD DATES AND BASIC PENETRATION RATES FOR LARGE-CITY SYSTEMS

NO.	CITY	STATE	AWARD	BEGAN	HP	BSUBS	BPEN
136	Long Beach	CA	1964	1965	145000	40000	0.28
137	Macon	GA	1964	1964	54278	34693	0.64
138	San Francisco	CA	1964	1964	218500	88801	0.41
139	Waco	TX	1964	1965	65202	35849	0.55
140	Austin	TX	1963	1963	187000	105000	0.56
141	Fayetteville	NC	1963	1964	82800	47800	0.58
142	Melbourne	FL	1963	1964	99000	72500	0.73
143	New Haven	CN	1963	1964	94224	60445	0.64
144	San Diego-Cox	CA	1963	1963	406439	251892	0.62
145	Glendale	CA	1962	1962	100418	51085	0.51
146	Santa Barbara	CA	1961	1962	73000	61778	0.85
147	Huntsville	AL	1954	1954	58000	35952	0.62
148	Eugene	OR	1953	1954	78500	44471	0.57
149	Reno	NV	1953	1953	68783	35816	0.52
150	Monterey	CA	1952	1952	80000	63000	0.79
151	Seattle-GW Center	WA	1952	1952	225000	104000	0.46
152	Seattle-Viacom	WA	1952	1952	160200	83000	0.52
	AVERAGE				18915113	9412338	0.50

SOURCES: CableFile (1986) and Television and Cable Factbook (1986).
NOTE: Data for homes passed (HP), basic subscribers (BSUBS), and basic penetration (BPEN) are for 1986.

APPENDIX D
BASIC PENETRATION RATES IN 1977 FOR LARGE-CITY SYSTEMS AWARDED PRIOR TO 1974

NO.	CITY/STATE		AWARD	BEGAN	HP	BSUBS	BPEN
1	Manhattan	NY	1970	1971	348000	90000	0.26
2	Los Angeles-GW	CA	1968	1969	256833	80743	0.31
3	San Diego-Cox	CA	1963	1963	230000	129750	0.56
4	San Jose	CA	1967	1968	166000	63000	0.38
5	San Francisco	CA	1964	1964	150477	46428	0.31
6	Buffalo	NY	1966	1966	140000	25000	0.18
7	Akron	OH	1968	1970	118300	18948	0.16
8	Toledo	OH	1966	1966	106000	50000	0.47
9	Oakland	CA	1969	1970	104567	15338	0.15
10	Saint Petersburg	FL	1972	1972	101902	16753	0.16
11	Columbus-W	OH	1970	1971	98500	27267	0.28
12	Seattle-GW Center	WA	1952	1952	97987	28891	0.29
13	Honolulu	HA	1973	1973	90000	27300	0.30
14	Rockford	IL	1973	1973	86500	49000	0.57
15	Fayetteville	NC	1963	1964	82800	47800	0.58
16	El Paso	TX	1968	1972	82748	26647	0.32
17	Austin	TX	1963	1963	80700	65087	0.81
18	Scranton	PA	1972	1973	75684	30281	0.40
19	Wilmington	DE	1968	1970	75000	46000	0.61
20	Orlando	FL	1970	1971	69900	25800	0.37
21	Seattle-Viacom	WA	1952	1952	67767	25519	0.38
22	Harrisburg	PA	1965	1965	66252	40448	0.61
23	Santa Barbara	CA	1961	1962	63338	42042	0.66
24	Corpus Christi	TX	1971	1972	62309	14549	0.23
25	San Diego-SW	CA	1965	1965	62100	25700	0.41
26	Colorado Springs	CO	1967	1969	62000	21600	0.35
27	Madison	WI	1965	1973	60000	20500	0.34
28	Columbus-C	OH	1973	1973	60000	25500	0.43
29	Lincoln	NE	1967	1968	58612	27258	0.47
30	W. Palm Peach	FL	1972	1972	58246	15868	0.27
31	Albany	NY	1968	1974	56300	23400	0.42
32	Eugene	OR	1953	1954	54416	33621	0.62
33	Overland Park	KA	1971	1971	53000	16300	0.31
34	Pompano Beach	FL	1969	1969	52553	7249	0.14
35	Columbus-ATC	OH	1971	1973	50400	14300	0.28
36	Peoria	IL	1966	1973	50000	23500	0.47
	AVERAGE				97200	35761	0.37

SOURCE: Television and Cable Factbook (1978).

NOTE: A large city system is defined as a system having at least 50,000 homes passed (HP) in 1977. BSUBS is the number of basic service subscribers in 1977. BPEN is basic service penetration.

APPENDIX E

AN ALTERNATIVE MEASURE OF CROSS-SECTIONAL REPUTATION

EXPLANATORY VARIABLES	
CONSTANT	0.136 (0.68)
BIDDERS	-0.010 (-1.37)
IDIOSYNK	-0.003** (-2.70)
NONPROF	-0.663** (-4.80)
HOTELS	1.114** (5.13)
MDU	1.093* (1.87)
REGULARTV	0.462** (5.06)
REPUTATIONALT	0.097 (0.06)
YRSOP	-0.326** (-3.48)
OUTLIER	9.416** (2.95)
RSQ	0.56
CRRSQ	0.48

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: Dependent variable in the ordinary least-squares regression is OVEREST. T-statistics are in parentheses.

APPENDIX F

EXPLAINING BASIC PENETRATION MISTAKES WITH NO ACCOUNT BEING MADE FOR THE OUTLIER

EXPLANATORY VARIABLES	
CONSTANT	0.244 (1.16)
BIDDERS	-0.009 (-1.05)
IDIOSYNK	-0.003** (-2.45)
NONPROF	-0.638** (-4.25)
HOTELS	0.920** (3.98)
MDU	1.063 (1.65)
REGULARTV	0.320** (3.65)
REPUTATION	0.013 (0.10)
YRSOP	-0.280** (-2.78)
RSQ	0.46
CRRSQ	0.38

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: Dependent variable in the ordinary least-squares regression is OVEREST. T-statistics are in parentheses.

APPENDIX G

EXPLAINING PAY PENETRATION MISTAKES

EXPLANATORY VARIABLES	
CONSTANT	-0.433 (-1.36)
BIDDERS	-0.014 (-1.14)
IDIOSYNK	0.001 (0.24)
NONPROF	-0.165 (-0.74)
HOTELS	0.779* (2.21)
MDU	1.461 (1.52)
REGULARTV	0.408** (2.80)
REPUTATION	-0.162 (-0.75)
YRSOP	-0.040 (-0.27)
OUTLIER	1.038* (1.90)
RSQ	0.20
CRRSQ	0.07

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: Dependent variable in the ordinary least-squares regression is POVEREST. T-statistics are in parentheses.

BIBLIOGRAPHY

Akerlof, George A., "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," Quarterly Journal of Economics, August 1970, Vol. 84, No. 3, pp. 488-500.

Axelrod, Robert, The Evolution of Cooperation, New York: Basic Books, Inc., 1984.

Cable Television Information Center, CableReports, Washington, D.C.: CTIC, monthly, various issues.

Besen, Stanley E. and Robert W. Crandall, "The Deregulation of Cable Television," Journal of Law and Contemporary Problems, Winter 1981, Vol. 44, No. 1, pp. 77-124.

Demsetz, Harold E., "Why Regulate Public Utilities?," Journal of Law and Economics, April 1968, Vol. 11, No. 1, pp. 55-65.

Dubin, Jeffrey A. and Peter Navarro, "How Markets for Impure Public Goods Organize: The Case of Household Refuse Collection," unpublished Social Science Working Paper 633, California Institute of Technology, March 1987.

Fairchild Business Publications, Multichannel News, Denver: Fairchild Business Publications, various issues.

Fudenberg, Drew and Jean Tirole, "A 'Signal-Jamming' Theory of Predation," Rand Journal of Economics, Autumn 1986, Vol. 17, No. 3, pp. 366-376.

Goldberg, Victor P., "Regulation and Administered Contracts," Bell Journal of Economics, Autumn 1976, Vol. 7, No. 2, pp. 426-448.

International Thomson Communications, CableFile, Denver: International Thomson Communications, yearly, various years.

----------, CableVision, Denver: International Thomson Communications, weekly, various weeks.

Joskow, Paul L., "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants," Journal of Law, Economics, and Organization, Spring 1985, Vol. 1, No. 1, pp. 33-80.

Kagan Associates, Inc., Cable TV Franchising, Carmel, California: Paul Kagan Associates, Inc., weekly, various weeks.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," American Economic Review, September 1986, Vol. 76., No. 4, pp. 728-741.

Klein, Benjamin, Robert Crawford, and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, October 1978, Vol. 21, No. 3, pp. 297-326.

----------, and Roy W. Kenney, "Contractual Flexibility," unpublished paper, revised January 10, 1986.

----------, and Keith Leffler, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, December 1981, Vol. 89, No. 4, pp. 615-641.

Lewis, Tracy R., "Reputation and Contractual Performance in Long-Term Projects," *Rand Journal of Economics*, Summer 1986, Vol. 17, No. 2, pp. 141-157.

MacAvoy, Paul W., ed., *Deregulation of Cable Television*, Washington, D.C.: American Enterprise Institute, 1977.

Macauley, Stewart, "Non-contractual Relations in Business," *American Sociological Review*, February 1963, Vol. 28, No. 1, pp. 55-67.

Masten, Scott E., "The Organization of Production: Evidence from the Aerospace Industry," *Journal of Law and Economics*, October 1984, Vol. 27, No. 2, pp. 403-417.

Milgrom, Paul and John Roberts, "Predation, Reputation and Entry Deterrence," *Journal of Economic Theory*, August 1982, Vol. 27, No. 2, pp. 280-312.

Monteverde, Kirk and David J. Teece, "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics*, Spring 1982, Vol. 13, No. 1, pp. 206-213.

Muris, Timothy J., "Opportunistic Behavior and the Law of Contracts," *Minnesota Law Review*, April 1981, Vol. 65, No. 4, pp. 521-590.

Owen, Bruce M., "The Rise and Fall of Cable Television Regulation," in *Case Studies in Regulation, Revolution, and Reform*, Leonard W. Weiss and Michael W. Klass, eds., pp. 86-101, Boston: Little, Brown & Co., 1981.

Palay, Thomas M., "Comparative Institutional Economics: The Governance of Freight Rail Contracting," *Journal of Legal Studies*, June 1984, Vol. 13, No. 2, pp. 265-288.

Posner, Richard A., *Economic Analysis of Law*, 2nd. ed., Boston: Little, Brown and Company, 1986.

Prager, Robin Ann, "Firm Behavior in Franchise Monopoly Markets: The Case of Cable Television," unpublished Ph.D. dissertation, M.I.T., 1986.

Radner, Roy, "Competitive Equilibrium Under Uncertainty," Econometrica, January 1968, Vol. 36, No. 1, pp. 31-58.

Schelling, Thomas C., The Strategy of Conflict, Cambridge, Massachusetts: Harvard University Press, 1960.

Scherer, Frederic M., Industrial Market Structure and Economic Performance, 2nd ed., Boston: Houghton Mifflin Company, 1980.

Selten, Reinhard, "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory, January 1975, Vol. 4, No. 1, pp. 25-55.

Shew, William B., Costs of Cable Television Franchise Requirements, White Plains, New York: National Economic Research Associates, 1984.

Simon, Herbert A., "Theories of Decision-making in Economics and Behavioral Science, American Economic Review, June 1959, Vol. 49, No. 3, pp. 253-283.

----------, Administrative Behavior, 2nd ed., New York: The Macmillan Company, 1961.

----------, "Rationality as Product and Process of Thought," American Economic Review, May 1978, Vol. 68, No. 2, pp. 1-16.

Television Digest, Inc., Television and Cable Factbook, Washington, D.C.: Television Digest, Inc., yearly, various years.

----------, Weekly Cable Action Update, Washington, D.C.: Television Digest, Inc., weekly, various weeks.

Webb, G. Kent, The Economics of Cable Television, Lexington, Massachusetts: Lexington Books, 1983.

Williamson, Oliver, "Franchise Bidding for Natural Monopolies--in General and with Respect to CATV," Bell Journal of Economics, Spring 1976, Vol. 7, No. 1, pp. 73-104.

----------, The Economic Institutions of Capitalism, New York: Free Press, 1985.

Zupan, Mark A., "Franchise Bidding for Natural Monopolies: How well does it Promote Efficiency in Local Cable TV Markets?," unpublished working paper, USC School of Business, Department of Finance and Business Economics, March 1987a.

----------, "Cable Franchise Renewals: Do Incumbent Firms Behave Opportunistically?," USC School of Business, Department of Finance and Business Economics, October 1987b.

CABLE FRANCHISE RENEWALS:
DO INCUMBENT FIRMS BEHAVE OPPORTUNISTICALLY?

by

Mark A. Zupan*

Department of Finance and Business Economics
USC School of Business
Los Angeles, CA 90089-1421
(213) 743-5778

Revised, April 1988

*The author would like to thank two anonymous referees, Tim Bresnahan, David Butz, Richard Craswell, Harold Demsetz, Art DeVany, Tom Gilligan, Ami Glazer, Doug Joines, Paul Joskow, Joseph Kalt, Benjamin Klein, Michael Levine, Tony Marino, Richard Posner, Richard Schmalensee, Dan Spulber and Clas Wihlborg and workshop participants at the University of Arizona, USC and UCLA for helpful comments and suggestions. Elden Chang, David Franzblau, Chris Murdoch and William Stevens provided research assistance.

I. INTRODUCTION

A recent systematic study of cable television franchises (Zupan, 1987) suggests that, with regards to contractual fulfillment, cable operators are typically well-behaved. Reneging on franchise commitments is infrequent and when it occurs it appears to be due largely to unforeseen changes in demand or cost.

The finding that franchise promises generally are kept, however, does not necessarily imply the absence of opportunistic behavior by cable operators. Since franchise contracts are not everlasting and are typically subject to renewal every 15 years, there is an important alternative manner in which an operator may rely on the idiosyncratic investments associated with a franchise contract to "hold up" (Victor Goldberg, 1976) a city for favorable changes in the contractual relationship. Specifically, operators may adhere to their promises during the initial contract period, but parlay the advantages of incumbency into concessions from the city at the time of franchise renewal.

Previous analyses (Alan Peacock and Charles Rowley, 1972; Richard Schmalensee, 1979; and Oliver Williamson, 1976) point to the lack of bidding parity during contract renewal as a central shortcoming of franchise bidding schemes. Schmalensee (1979, pp. 72-73), for example, advances several *a priori* arguments as to why effective competition is unlikely to obtain at the time of refranchising:

> In the first place, it seems virtually impossible to ensure that outsiders would pay for the tangible assets exactly what they are worth to the existing supplier in their current use. Any discrepancy between the transfer price and value to the existing supplier creates an asymmetry in bidding that is likely to favor the current supplier. At the very least, the current supplier avoids the haggling costs associated with asset transfer if he wins. Second, the existing supplier is almost certain to be better informed about actual cost and demand conditions

> than his rivals are. He thus saves the cost of acquiring the information necessary to submit an intelligent bid, and he may be able to mislead or hinder his rivals. Third, the severe problems associated with pricing (or transferring) skills and knowledge that are specific to a single enterprise raise bidding parity issues...Finally, any firm that managed to outbid an existing supplier would have to expect to go through a high-cost transition period during which its management team learned the firm-specific technical and personal aspects of their jobs. These inevitable costs associated with managerial change give the existing supplier yet another advantage at bidding.

Because of such a priori reasons, previous studies conclude that incumbent firms have "substantial discretionary power" (Peacock and Rowley, 1972, p. 242), opportunism will fail to be curbed (Williamson, 1976), and "franchise bidding is hardly a breakthrough in natural monopoly control technology" (Schmalensee, 1979, p. 73).

Although incumbent cable firms may have considerable advantages over potential entrants, the extent to which such advantages foster opportunistic conduct at renewal time has never been empirically investigated. Nor has mention been made of certain important theoretical reasons for expecting that bad behavior by incumbent operators will be the exception rather than the rule. Among these reasons are reputational constraints (Benjamin Klein and Keith Leffler, 1981) and the potential for vertical integration by cities into the local distribution of CATV services (Paul Joskow, 1985; and Klein, Robert Crawford, and Armen Alchian, 1978). The ability of operators to exploit the advantages of incumbency is also circumscribed by the countervailing power of cities to do likewise; i.e., while their irreversible investments place existing firms in a superior position vis-a-vis potential entrants, the sunk costs have value only to the city in which they are sunk. The monopsony power the incumbent city can bring to bear on the incumbent operator at renewal time may be quite substantial--especially since many cities have a say in determining the "fair value" at which the assets of an operator failing to obtain renewal may be sold for.

This paper attempts to remedy the shortcomings of the existing literature in two ways. First, the a priori arguments which suggest that badly behaving firms will be a relatively rare species at the time of refranchising are more fully advanced and discussed. Second, a systematic empirical investigation is undertaken of recent cable refranchisings. The goals of the investigation are to uncover the extent and nature of any operator opportunism that may be present at refranchising time.

II. TOWARDS A BROADER AND CLEARER DEFINITION OF OPPORTUNISM

While opportunistic behavior is certainly a hazard of contracting (cf., Williamson, 1985; and Joskow, 1985), it is more generally a peril associated with economic relationships. Aside from the case of once-for-all accords such as marriage, there is no necessary one-to-one correspondence between the duration of a contract and the length of time over which a relationship proves valuable to the parties striking the contract.[1] Indeed, with recurrent, short-term transactions involving specialized labor, plant, equipment, technology and/or organizational structure, a relationship often spans a sequence of contracts.

Where a relationship consists of a series of contracts, opportunism may seriously infect the relationship even though each and every contract in the relationship evidences immunity from insidiousness. Imagine, for instance, that marriage was a 5-year-with-option-for-renewal agreement rather than a once-for-life covenant. We may well observe husband and wife dutifully fulfilling their respective obligations during every 5-year accord and yet still witness one spouse opportunistically leveraging the irrevocable investments associated with the relationship into favorable contractual changes every time the agreement is renewed.

Any investigation for opportunism in an economic relationship--whether such an investigation focuses on the terms of the economic relationship during a contractual period or at the time of a recontracting--must, of course, be predicated on a definition of opportunism that is both theoretically precise and capable of being empirically implemented. What exactly are we looking for when we look for opportunistic behavior?

While its easy to find economists who view opportunism as being a fundamental problem associated with economic relationships, finding a clear and workable definition of opportunism in the economics literature is a much more difficult task. The definitions that have typically been posited in the literature tend to be theoretically fuzzy and hard to empirically implement. Williamson (1985, pp. 47-49), for example, defines opportunism as being:

> self-interest seeking with guile. This includes but is scarcely limited to more blatant forms, such as lying, stealing, and cheating. Opportunism more often involves subtle forms of deceit...More generally, opportunism refers to the incomplete or distorted disclosure of information, especially to calculated efforts to mislead, distort, disguise, obfuscate, or otherwise confuse.

Joskow (1985, p. 37) takes opportunism to refer to both "behavior that does not maximize joint profits (and is inefficient) when a particular contingency arises and also behavior that involves the appropriation of wealth of one party by the other in some states without necessarily inducing distortions in supply and demand."

To more precisely pin down the concept of opportunism and to make the idea operational, it would seem more appropriate to focus on an economic relationship's terms of trade and on how these terms change, if at all, after the parties involved make specialized investments in the relationship, *ceteris* paribus. Should the terms of trade be altered significantly in the buyer's favor once the specialized investments are made, opportunism on the part of the buyer would appear to be afoot. Analogously, the seller could be deemed to have behaved opportunistically should the terms of trade be altered significantly in the seller's favor.[2]

For an example, consider the typical agreement between a cable operator-seller and a city-buyer. The terms of trade generally involve, although never in complete contingent claims fashion, a construction deadline, basic and pay tier programming and rates, and certain nonprice concessions. The basic tier bundles together local broadcast stations and satellite- or microwave-fed services such as MTV, ESPN and CNN. Pay tiers provide, largely in unbundled fashion, higher-value, satellite-fed entertainment services such as HBO and Showtime. The nonprice concessions are largely fixed cost in nature and may include: direct endowments; free hookups for public institutions; community programming; institutional networks (I-nets) linking various public facilities in a city; excess channel capacity; and a franchise fee levied as a percentage of operating revenues.

In the case of basic tier pricing, suppose that for given city preferences and operator technology, the rate that the city and operator may potentially agree on prior to initiating their relationship ranges anywhere from Pb (the rate most favorable to the operator--i.e., the monopoly rate) to Ps (the rate that is most favorable to the city but still sufficient to allow the operator to cover costs).[3] And suppose that the operator and city agree to an intermediate rate, Pnoii (where Pb > Pnoii > Ps), prior to making any idiosyncratic investments in their relationship. We may then conclude that problems of opportunism do not infect the relationship if the price that we observe after the idiosyncratic investments are made, Pii, does not differ from the price agreed to prior to the making of the investments, Pnoii. In other words, assuming that all else has been held constant, the case of no opportunism is defined by:

$$Pnoii = Pii \qquad \text{(NO OPPORTUNISM).} \qquad [1]$$

In contrast, operator-seller opportunism may be presumed where the price observed after the idiosyncratic investments are made exceeds the price agreed to ex ante:

$$Pnoii < Pii \qquad \text{(OPERATOR-SELLER OPPORTUNISM);} \qquad [2]$$

while the converse relationship implies city-buyer opportunism:

$$Pnoii > Pii \qquad \text{(CITY-BUYER OPPORTUNISM)} \qquad [3]$$

The foregoing suggests a straightforward test for the presence of operator opportunism at the time of cable franchise renewal. Suppose that, on a cross-sectional basis, we compare the terms of trade in renewal contracts with the terms of trade in concurrently-struck initial franchise agreements. If everything else is held constant, we may safely conclude that operator opportunism is present only if the terms of trade are significantly more favorable to operators in cities renewing their franchise contracts than they are in cities concurrently-issuing initial franchise awards.

III. THE CONSTRAINTS ON OPERATOR OPPORTUNISM

Before testing for the presence of operator opportunism at the time of cable refranchising, it is important to note the constraints which serve to inhibit bad behavior on the part of cable operators. While previous authors (cf., Williamson, 1976) have argued that operators may be tempted to exploit the quasi-rents associated with the typical franchise relationship, there are at least four important mechanisms which serve to restrain operators from behaving badly at refranchising time. These restraints include:

1. vertical integration;
2. cross-sectional reputation--a badly behaving operator may be unable to attract new customers if existing customers are maltreated;
3. reputation internal to an existing relationship--bad behavior by cable operators may induce retaliation and lower-than-optimal investment in the relationship by current customers; and
4. monopsony power possessed by current customers--the irrevocable investments made by cable operators have value only in the cities in which they are sunk and these values are thus subject to expropriation by those cities.

III.A. Vertical Integration

Vertical integration provides one potential limit on operator opportunism. Where prospective partners foresee that a relationship will require the investment of nonsalvageable assets and that the quasi-rents associated with these assets may be ripe for appropriation (either during a contract or at contract renewal time), vertical integration may prove more desirable than the striking of a renewable, incomplete contract (Klein, Crawford, and Alchian, 1978). Systematic empirical evidence supporting this hypothesis continues to amass. From the automobile industry (Kirk Monteverde and David Teece, 1982), the organization of production in the aerospace industry (Scott Masten, 1984), supply arrangements between coal producers and electric utilities (Joskow, 1985), and the provision freight-rail services (Thomas Palay, 1984), vertical integration appears to be more prevalent where the risks of opportunistic behavior are greater.

In local cable markets, vertical integration takes the form of municipal ownership: the proprietorship and operation of the system by the city. Perhaps tellingly, less than 1 percent of all cable systems in the U.S. are municipally owned and only .1 percent of all subscribers are served by such systems.[4]

III.B. Cross-Sectional Reputation

Besides vertical integration, reputation may also serve to promote upstanding conduct. Previous authors (cf., Klein and Leffler, 1981) have correctly pointed to the lost business that notorious actions may entail. Importantly, future business may be lost either from existing or from new customers.

At least in the case of local CATV franchises, the threat of lost business from new customers (i.e., cross-sectional reputation) may be quite effective at curbing operator opportunism. Of the 36 million cable subscribers in 1985, well over 90 percent were accounted for by multiple system operators (MSOs).[5] To the extent that cities place stock in a cable firm's performance elsewhere and that an operator intends to win future franchise bidding competitions held by other cities and/or assure local policymaker approval of prospective system purchases, abstention from opportunistic conduct both during a contract and at refranchising time becomes desirable.[6]

III.C. Reputation Internal to the Existing Relationship

In addition to being valued cross-sectionally, a reputation for virtue may be highly prized within an existing relationship. Where the final product's worth depends on both party's effort levels over the course of the relationship, brazenly self-serving conduct may provoke similar treatment from the other side in retaliation and leave both sides with a lower-valued final product (Robert Axelrod, 1984; Tracy Lewis, 1986; and Stewart Macauley, 1963). Indeed, there are many ways besides initiating litigation that a city could make life unpleasant for a cable operator in response to a hold up, be it 4 years into a 15-year contract or at contract renewal time. The operator may be called before the city council more frequently. City policymakers may impose property tax increases, hike utility pole rental rates or drag their feet on granting easements. Dogged persistence may be displayed to ensure that the operator adheres to the fine print of the contract, no matter how small the fine print.

Virtue proves valuable within an ongoing relationship for the same reasons that: the maximization of joint benefits is more likely when a rivalry is continuous (Axelrod, 1984; Thomas Schelling, 1960; and Frederic Scherer, 1980); and cooperation may emerge between competitors who, at any point in time, compete in numerous markets (Corwin Edwards, 1955; and Scherer, 1980). According to Scherer (1980, p. 163), a dynamic rivalry fosters a quid pro quo relationship and mutually beneficial behavior since "Repeated experience under stable conditions affords an opportunity for learning to cooperate and trust one another; and when the game will be played continuously or repeatedly, each party can threaten its rivals with damaging retaliation tomorrow if cooperation is not forthcoming today."[7]

III.D. Cities' Monopsony Power

Previous analyses (cf., Williamson, 1976, p. 83 and p. 89) argue that the irrevocable commitments made by operators "place winners of the original competition at a substantial advantage over nonwinners" and convert "what was once a large numbers bidding situation, at the time the original franchise was awarded...into what is tantamount to a small numbers bargaining situation when the franchise comes up for renewal." What the prior studies have failed to emphasize, however, is that incumbent cable firms have in a sense earned these nontrivial advantages; resources have been expended by the firms to acquire bidding superiority at refranchising time. The ability of operators to exploit the advantages of incumbency, furthermore, is constrained by the fact that an operator's nonsalvageable investments have value to only one buyer: the city in which the investments have been sunk.

While prior research has typically pointed to the monopoly power opportunistic operators may exercise, the counteracting monopsony power a city can exert may be substantial. And cities may be as capable as operators of behaving badly.[8] Stephen Effros, president of the trade group Community Antenna Television Association, indeed argues:[9]

> Renewal really isn't the problem; its the cost of renewal that's the problem. Its the price an operator has to pay to get the renewed franchise. It's clear that an operator is not going to allow an asset that he's developed over 15 years to simply vanish because a city decides on a whim that it will now award the franchise to someone else. Nevertheless, because the operator is forced to protect that asset every 15 years, the city has the opportunity to say to the cable company, 'You must give us this and that to get the renewal.' The operator with his assets already in the ground, has no choice except to say o.k.

The monopsony power possessed by incumbent cities is amplified by the fact that a sizeable number of franchise contracts give cities a say in determining the "fair value" at which the assets of an operator failing to obtain renewal may be sold.[10] Where cities have at least a partial ability to authoritatively specify the worth of an operator's assets, the operator's bargaining strength at renewal time is diminished.[11] The incentive for firms to behave insidiously hence is lessened; put colloquially, it rarely pays to antagonize those who have you by the short hairs.

IV. EMPIRICAL EVIDENCE

To what extent are the preceding constraints successful at deterring operator opportunism at the time of franchise renewal? The discussion in Section II suggests that we should compare the terms of trade in renewal contracts with the terms of trade in concurrently-struck initial franchise contracts. If there are significant idiosyncratic investments associated with cable franchise relationships and if the defense mechanisms preventing operators from opportunistically exploiting such idiosyncratic investments are not well-working, most cable contracts should be renewed and the terms of trade should be more favorable to operators in renewal contracts than in concurrently-struck initial franchise contracts, ceteris paribus.[12]

IV.A. The Frequency of Franchise Renewal

Based on the assumption that franchise contracts average 15 years in length (Zupan, 1988), Table 1 provides a schedule of the number of franchises expiring between the years 1967 and 2001. Of the approximately 7600 franchised systems in existence in 1986, 37 percent (2841 systems) have undergone refranchising at least once and 7 percent (500 systems) have been subject to renewal twice. While the current annual number of refranchisings is already quite large (an average of 268 systems per year over the period 1980-87), this number will nearly quadruple by the turn of the century.

Prior studies (Peacock and Rowley, 1972; Schmalensee, 1979; and Williamson, 1976) imply that the idiosyncratic investments associated with cable franchises are sizable enough to ensure franchise renewal. This

TABLE 1

A CABLE REFRANCHISING SCHEDULE

Year	Initial Refranchisings	Second/Third Refranchisings	Total Refranchisings
1967	70	0	70
1968	80	0	80
1969	150	0	150
1970	100	0	100
1971	50	0	50
1972	50	0	50
1973	25	0	25
1974	35	0	35
1975	80	0	80
1976	60	0	60
1977	100	0	100
1978	200	0	200
1979	200	0	200
1980	125	0	125
1981	245	0	245
1982	200	70	270
1983	230	80	310
1984	260	150	410
1985	230	100	330
1986	149	50	199
1987	202	50	252
1988	150	25	175
1989	167	35	202
1990	348	80	428
1991	175	60	235
1992	151	100	251
1993	43	200	243
1994	275	100	375
1995	75	125	200
1996	150	245	395
1997	450	270	720
1998	775	310	1085
1999	600	410	1010
2000	400	330	730
2001	1000	199	1199

SOURCE: *Television and Cable Factbook* (1986).
NOTE: The schedule is based on the assumption that the duration of all cable franchises is 15 years and that all refranchising is conducted in the final year of existing contracts.

implication is resoundingly confirmed by historical experience. Of the over 3300 refranchising decisions made by cities to date, only the 5 cities listed in Table 2 have kicked out an incumbent operator.[13]

It is the general consensus in the industry that nonrenewal is nondesirable. The leading consulting firms CTIC and Malarkey, Taylor and Associates, for example, both state that their first recommendation to client cities is to negotiate with the existing firm. Explains CTIC executive director Harold Horn: "We start from the premise that the city is best off to reconcile itself with its existing operator even if there are problems."[14] Operators display an analogous preference for extending relationships with incumbent cities; at least partly from the fear of the diminution in the value of operator assets that may result from nonrenewal.[15]

Nonrenewal occurs only where the irrevocable investments associated with a franchise relationship are minor. Factors which lower the level of operator-city commitment include:

1. a small channel capacity;
2. obsolescent cable plant requiring a rebuild;
3. a need for additional channel capacity--provided that the desired capacity can be added most economically by rebuilding from scratch rather than by upgrading the existing plant;
4. a minimal amount of city-operator interrelationship through features such as community programming;
5. a turnover in cable proprietorship resulting from a system sale by the original franchisee;
6. demographic, economic and ideological changes in a city's population or in the policymakers representing that population.

All five of the incumbents failing to obtain renewal have been senescent 12-channel systems (capacities range from 12 to 140 channels) in need of both

TABLE 2

CITIES WHICH HAVE DENIED RENEWAL TO AN INCUMBENT CABLE OPERATOR

1. Danville, Kentucky
2. Marquette, Michigan
3. Morganton, North Carolina
4. Santa Cruz, California
5. Valparaiso, Florida

SOURCE: Cable TV Franchising, January 31, 1987, p. 6.

a rebuild and a channel capacity upgrade. None of the five featured extensive levels of community programming. Three of the five (Marquette, Santa Cruz and Morganton) obtained the franchise through a system sale by the original proprietor. In Santa Cruz, refranchising problems began in 1984, only 3 years after the system had been purchased by Group W from Teleprompter.[16]

IV.B. A Comparison of Renewal Contracts and Concurrently-Struck Initial Franchise Agreements

While the foregoing evidence suggests that, as far as refranchising is concerned, it pays to stick with one's existing contractual partner, the question remains whether renewal contracts evidence opportunism on the part of cable operators. In an attempt to provide an answer to this question, the terms of trade in two different types of contracts are examined: [i] a sample of 59 renewal accords agreed upon over the period 1980-1984 for which some relevant contractual (as well as cost and demand) data are available from the weekly industry newsletter, Cable TV Franchising; and [ii] a random sample of 66 concurrently-struck initial franchise awards for which detailed demand, cost and contractual data were obtained through a telephone survey conducted in the fall of 1984 (Zupan, 1988).[17] The terms of trade which are compared include channel capacity, franchise fee, dedicated community programming channels, basic tier price, basic tier price per channel of service offered and lead pay tier price.[18] Table 3 provides some basic statistics on these terms of trade for the two samples.[19]

IV.B.1. Promised Channel Capacity, Franchise Fee and Channels Dedicated to Community Programming

The channel capacity (CHCAP), franchise fee (FFEE) and number of community programming channels (COMCH) promised by an operator at the time of

TABLE 3

SOME BASIC STATISTICS ON CABLE FRANCHISE TERMS OF TRADE: RENEWAL CONTRACTS VS. CONCURRENTLY-STRUCK INITIAL CONTRACTS

	Initial Contracts	Renewal Contracts
Channel Capacity		
Average	46.4	46.8
Standard Deviation	20.7	11.6
Number of Observations	66	59
Franchise Fee (as a Percent of Total Revenue)		
Average	2.9	3.6
Standard Deviation	1.7	1.2
Number of Observations	47	48
Dedicated Community Programming Channels		
Average	2.8	3.4
Standard Deviation	3.6	2.4
Number of Observations	66	58
Monthly Basic Tier Price (as of 10/84)		
Average	$9.35	$8.86
Standard Deviation	$1.84	$1.64
Number of Observations	66	23
Monthly Basic Tier Price Per Channel of Programming Offered (as of 10/84)		
Average	$0.52	$0.51
Standard Deviation	$0.30	$0.18
Number of Observations	66	23
Monthly Lead Pay Tier Price (as of 10/84)		
Average	$9.51	$10.63
Standard Deviation	$1.27	$1.36
Number of Observations	66	23

SOURCES: See text.
NOTES: Statistics regarding franchise fees exclude data from systems located in the state of Massachusetts. See text for details.

franchising/refranchising should depend on exogenous factors such as the: size of the community (LOGHOMES);[20] per capita income level in the community (INCOME);[21] number of over-the-air broadcast signals available to the community (MUSTCARRY);[22] and year of contract agreement (YEAR).[23] Table 4 reports the OLS results that obtain when CHCAP, COMCH and FFEE are individually regressed on a set of explanatory variables which account for such exogenous factors as well as a dummy variable (RENEW) to account for whether the terms of trade derive from a renewal contract or not.[24]

As can be seen, only in the case of channel capacity are the terms of trade significantly more favorable to cable operators in renewal contracts than in concurrently-struck initial franchise awards. The channel capacity result, moreover, is due to the fact that three of the systems obtaining initial franchise awards promised economically-wasteful, dual cable systems with at least 108 channels (e.g., two 54-channel cables of which only one is used).[25] In contrast, none of the systems obtaining a renewal contract promised more than a single cable with 70 channels. When an explanatory variable is introduced to account for the three dual cable, excess capacity systems in the initial award sample (DUALEXCESS), there is no statistically significant difference between renewal and initial contracts in terms of promised channel capacity--as shown in Table 5.[26]

IV.B.2. Basic and Lead Pay Tier Prices

In reduced form, the monthly basic tier price (BPRICE), basic tier price per channel of service offered (BPPCH), and lead pay tier price (PPRICE) charged by an operator should depend on predetermined demand, cost and institutional factors.[27] To further investigate, therefore, whether operators behave opportunistically at refranchising time, the levels of BPRICE, BPPCH

TABLE 4

AN ECONOMETRIC ANALYSIS OF CABLE FRANCHISE CONTRACT TERMS OF TRADE: CHANNEL CAPACITY, FRANCHISE FEE AND THE NUMBER OF CHANNELS DEDICATED TO COMMUNITY PROGRAMMING

EXPLANATORY VARIABLE	CHANNEL CAPACITY	FRANCHISE FEE	COMMUNITY CHANNELS
LOGHOMES	5.33** (5.32)	0.30** (2.85)	1.15** (6.31)
INCOME	1.40* (2.03)	0.10 (1.22)	0.21* (1.69)
MUSTCARRY	0.74* (2.14)	0.02 (0.35)	0.06 (0.88)
YEAR	3.00** (2.42)	0.04 (0.29)	0.28 (1.23)
RENEW	-8.99** (-2.46)	0.22 (0.54)	-0.84 (-1.27)
CONSTANT	-22.59** (-2.56)	-0.62 (-0.69)	-9.97** (-6.21)
RSQ	0.37	0.20	0.37
CRRSQ	0.34	0.15	0.34
TOTAL OBS.	125	95	124
RENEWAL OBS.	59	48	58

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: T-statistics are in parentheses.

TABLE 5

ACCOUNTING FOR THE THREE DUAL CABLE, EXCESS CAPACITY SYSTEMS IN THE SAMPLE OF CITIES AWARDING INITIAL FRANCHISE CONTRACTS

EXPLANATORY VARIABLE	CHANNEL CAPACITY	FRANCHISE FEE	COMMUNITY CHANNELS
LOGHOMES	3.92** (4.72)	0.28** (2.65)	0.99** (5.72)
INCOME	0.88 (1.56)	0.08 (0.91)	0.14 (1.23)
MUSTCARRY	0.82** (2.96)	0.02 (0.43)	0.07 (1.15)
YEAR	1.90* (1.88)	0.02 (0.15)	0.19 (0.88)
DUALEXCESS	0.99** (8.02)	0.02 (0.98)	0.12** (4.57)
RENEW	-3.16 (-1.04)	0.33 (0.77)	-0.18 (-0.28)
CONSTANT	-7.94 (-1.08)	-0.36 (-0.38)	-8.31** (-5.40)
RSQ	0.59	0.21	0.47
CRRSQ	0.57	0.15	0.44
TOTAL OBS.	125	95	124
RENEWAL OBS.	59	48	58

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: T-statistics are in parentheses.

and PPRICE in October 1984 for each system for which such data are available are individually regressed on: the basic penetration rate expected at the time of franchising/refranchising (EXPBPEN);[28] the size of the community (LOGHOMES);[29] a dummy variable reflecting the absence of price controls on basic rates (UNREGBASIC);[30] a dummy variable reflecting a lack of for-profit motivation on the part of an operator (NONPROFIT);[31] a dummy variable to account for any differences in regular basic tier and lead pay tier pricing in systems with an upper-level, expanded basic tier (EXPAND);[32] and the dummy variable RENEW.[33] The ordinary least squares results are shown in Table 6.[34]

As can be seen in Table 6, only in the case of lead pay tier pricing does there appear to be any possible evidence of operator opportunism at the time of refranchising. With regards to basic tier prices and the prices charged per channel of basic service, the terms of trade appear to be no more favorable to operators in renewal contracts than in initial franchise awards. The approximate one dollar difference in lead pay tier prices between renewal and concurrently-struck initial award systems, moreover, may not necessarily represent operator opportunism at the time of refranchising. Since pay rates cannot be subject to formal rate control according to federal law but nevertheless appear to be subject to some informal rate controls (Zupan, 1988), the one dollar difference may represent the gradual exercise of pay tier pricing power over the course of the initial franchise contract rather than operator opportunism at refranchising time.

V. SUMMARY AND CONCLUSION: THE GUYS WEARING THE BLACK HATS (IN THE ECONOMICS LITERATURE) ARE PREDOMINANTLY GOOD AND ONLY RARELY BAD AND UGLY

In a seminal case study of the Oakland cable franchising experience, Williamson (1976, p. 84) argues that operator opportunism will fail to be

TABLE 6

AN ECONOMETRIC ANALYSIS OF CABLE FRANCHISE CONTRACT TERMS OF TRADE: BASIC AND LEAD PAY TIER PRICES

EXPLANATORY VARIABLE	BASIC PRICE	BASIC PRICE PER CHANNEL OF SERVICE	LEAD PAY PRICE
LOGHOMES	-0.04	-0.10**	-0.21*
	(-0.34)	(-6.91)	(-2.30)
EXPBPEN	1.85	0.24	0.25
	(1.32)	(1.31)	(0.23)
UNREGBASIC	3.15**	0.22**	-0.10
	(4.55)	(2.48)	(-0.19)
NONPROFIT	-3.48**	-0.27	0.14
	(-2.61)	(-1.58)	(0.14)
EXPAND	-1.86**	0.13*	0.62
	(-3.57)	(1.96)	(1.51)
RENEW	-0.35	0.01	1.13**
	(-0.81)	(0.15)	(3.27)
CONSTANT	8.66**	1.28**	11.19**
	(6.34)	(7.30)	(10.35)
RSQ	0.25	0.45	0.21
CRRSQ	0.20	0.41	0.15
TOTAL OBS.	89	89	89
RENEWAL OBS.	23	23	23

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: T-statistics are in parentheses.

curbed because "the efficacy of recurrent short-term contracting depends crucially...on the assumption that parity among bidders at the contract renewal interval is realized." Williamson (1976, p. 83) provides powerful reasons why, in the context of cable television, "bidding parity at contract renewal will not obtain."

The assertion that bidding parity will be lacking at the time of recontracting is easily accepted and is indeed resoundingly confirmed by the empirical evidence presented in this paper. Out of over 3,300 cable refranchisings conducted to date, only 5 decisions have involved the replacement of an incumbent by a new entrant. The advantages of tenure appear to be significant enough to make nonrenewal substantially undesirable.

The assertion that incumbency advantages foster opportunistic behavior on the part of cable operators, however, is much more debatable and is the subject of systematic scrutiny by this and a companion piece (Zupan, 1987). The results of this scrutiny indicate that operators behave much better than is commonly believed. Reneging on promises during a contract is infrequent and when it occurs it appears to be due largely to unforeseen changes in cost and/or demand. Except for (possibly) leading to slightly higher pay tier prices, furthermore, the advantages of incumbency do not appear to be exploited by operators at the time of franchise renewal.

The defense mechanisms which appear to so effectively constrain opportunistic operator behavior are reputational considerations and the countervailing monopsony power possessed by cities. It is concerns over having a "bad name" and the incumbency advantages available to cities which elicit cooperative rather than chiseling behavior from cable operators.

FOOTNOTES

[1] Contract duration and the importance of relationship-specific investments well may be positively correlated, however (cf., Joskow, 1987).

[2] This assumes that everything else remains constant (e.g., that cost and demand conditions and the relevant parties' information sets are equivalent before and after the specialized investments are made).

[3] An operator's costs, of course, include any expenditures made on the nonprice concessions which may be requested by local policymakers.

[4] The most recent data on the extent of municipal ownership comes from the newsletter *Cable TV Franchising*, April 14, 1983, p. 6. While the absence of vertical integration in the provision of local cable service may indicate that operator opportunism is not a severe problem, it may also be due to other unrelated factors such as: learning-by-doing advantages possessed by private firms; the monopsony power multiple system private operators are able to exercise in input markets; and First Amendment and/or ideological considerations (cf., Jeffrey Dubin and Peter Navarro, 1987).

[5] The top 100 MSOs alone account for over 90 percent of the total subscribership (*CableFile*, 1986, p. I-5).

[6] Zupan (1987) documents the prominent role a bidder's track record occupies among the criteria cities rely upon to either accept or reject a bid. Even after all communities are franchised, cross-sectional reputation proves valuable provided that there are a nontrivial number of system sales and that cities look at firms' performance records when deciding whether or not to approve a sale. In 1986, system sales totaled roughly $8 billion and over 15 percent of all subscribers changed company hands (*CableVision*, February 16, 1987, p. 54).

[7] The importance of internal reputation in cable markets is noted by Roger Turner, president of Colony Communications, who argues that (*CableAge*, April 2, 1984, p. 19):

> when a franchise is won, that becomes the time for the MSO to become more intense in its communications with the community 'not only to promote itself, but more importantly to establish solid good will.'...[Colony] sends its director of governmental relations to each of its franchise cities on a regular quarterly basis to give first-hand presentations to franchise authorities...[the presentations] include what is new in the cable industry in terms of technology and services, what new programming services Colony is adding or why they are taking off a service, how the community can best utilize its public access channels, why rate increases become necessary, and an interface with authorities on what needs the communities may have...'In essence, we are trying to head off complaints and difficulties at the pass.'

[8] Opportunistic behavior by cities is constrained by, among other factors, internal and cross-sectional reputation. In an existing relationship, there are many ways in which an operator can retaliate against a badly-behaving city. Fewer outside firms will also be interested in a renegade city's franchise.

[9] CableAge, February 20, 1984, pp. 25-26.

[10] While the federal Cable Act of 1984 protects operators during the renewal process by curtailing the ability of cities to demand "exotic" services as a condition of renewal, it also enables all municipalities to buy back a system at "fair market value" if the franchise has been revoked or has expired (Cablevision, March 17, 1986, p. 54). Evidence on the prevalence of buy-back provisions prior to 1984 is based on telephone conversations with cable executives and analysts for consulting firms. Officials at the Cable Television Information Center (the leading consulting firm to cities), for instance, declined to give a percentage but stated that "numerous" contracts coming up for renewal before 1984 contained buy-back clauses. Tom Alexander, president of Alexander and Associates (a leading consulting firm to cable companies) estimated that 50 to 60 percent of all franchises contained buy-back clauses and that 75 to 80 percent of all big-city franchises contained such clauses. Andrew Holdgate, Director of Public Relations for Storer Communications (a major MSO), reported that 15 to 25 percent of their franchises included buy-back provisions.

[11] In a page-long footnote, Williamson (1976, p. 86) describes the provisions of a typical "buy-back" clause:

> in the event said franchise shall expire...said city [Los Angeles] shall have the right at its option...to purchase and take over the property of such utility, and in the event that said city shall so exercise its right...said city shall pay to the said grantee the fair value of the property...The price to be paid...shall be on the basis of actual cost to the utility for the property taken, less depreciation accrued, as of the date of purchase, with due allowance for obsolescence...no allowance will be made for franchise value, good will, going concern, earning power, increased cost of reproduction or increased value of right of way or allowance for damages by reasons of severance... the valuation...shall be determined by a board of three arbitrators of whom one shall be appointed by the city, one by its grantee, and the third by the two arbitrators so appointed...In the event the grantee shall fail to appoint an arbitrator...or in the event the two arbitrators appointed by the city and the grantee...shall fail to appoint a third arbitrator...such arbitrator shall be appointed by the presiding Judge of the Superior Court of the State of California.

[12] This assumes that the incumbency advantage is not entirely capitalized at the time of initial franchising--an assumption that is supported by the evidence in Robin Prager (1986) and Zupan (1987).

[13] CableVision, pp. 114-212. There have been cases, however, where dissatisfaction has prompted a sale prior to franchise expiration (cf., Cablevision, December 7, 1981, pp. 115-212).

[14] CableVision, December 7, 1981, p. 125. The CTIC newsletter advises cities that (CableReports, December 1980, p. 1):

> Generally speaking, the Center feels the franchisee should seriously be considered for renewal if the cable company has provided quality services, at reasonable rates, has kept its promises, and has complied with the terms and agreements made with the local government. Most likely, the city will find it has negotiating leverage with the existing operator who wants to keep the franchise...It is very likely that the existing operator, having an investment at stake, will be more responsive than a new company to community pressure to provide services in addition to those already offered on the old system...What the city must ask itself is whether the community has its best chance at a modernized system by keeping the same cable operator, and negotiating a new franchise. For many reasons this is often the case. The present cable operator's track record is known, the city and the company are well-acquainted, and can start a negotiation with the hard realities obvious to all...Furthermore, the community can avoid the possibility of a loss of service in the interim between shutting down an old operation and a new system being built.

[15] An industry newsletter points out that (Cable TV Franchising, May 31, 1983, p. 3):

> One reason cable operators are so concerned about refranchising is Cox Cable's sale of its Marquette, MI, area franchises to Marquette Cablevision, a local firm that successfully challenged Cox at renewal time. Settlement papers from the lawsuit show Marquette Cablevision purchased the assets of the Marquette franchise and 15 other franchises within 75 square miles of Marquette...Settlement documents show the systems generated $2.2 million in revenue and $1,073,000 in cash flow last year. Sale price: $4 million. That works out to 3.7xcash flow or $260/sub. Except for Munising, which is a 35-channel capacity system, all of the others are 12 channels. Although a rebuild is in order, the sale price is for only about 25-30% of fair market value.

[16] Multichannel News, February 25, 1985, p. 4 and December 2, 1985, p. 77.

[17] Cable TV Franchising is marketed by industry analyst Paul Kagan and furnishes information on political developments pertinent to local cable markets. Of the 59 renewal cases for which the newsletter provided some relevant data, 23 were detailed case studies based on interviews with city officials, consultants and cable operators. Since the case studies were selected so as to provide an accurate portrayal of the refranchising process,

they offer a fruitful means for examining whether operators behave opportunistically at refranchising time. Data for the remaining 36 renewal cases drawn from the Kagan newsletter, are generally less complete (particularly with regards to basic and lead pay tier prices and with regrads to whether or not a system was subject to basic rate controls in October 1984). Importantly, all econometric results of interest reported below that employ data from all 59 renewal cases do not differ significantly from the results that obtain when only the 23 more-finely-detailed renewal contracts are compared to the 66 concurrently-struck initial franchise awards. Wherever possible, the Kagan newsletter data were corroborated and supplemented by data obtained from the industry trade manuals Television and Cable Factbook and CableFile and, in certain cases, over the telephone from the individual cable system managers.

[18] The renewal data are not sufficient to permit a comparison of other key terms of trade such as I-net mileage, I-net channel capacity, spending commitments for community programming, and the number of channels that are both dedicated and used for community programming.

[19] The Table 3 franchise fee statistics exclude systems located in the state of Massachusetts since Massachusetts state law specifies that a cable system must pay $0.80 per year per basic subscriber to the state and $0.50 per year per basic subscriber to the local community. When the Massachusetts systems are included in the econometric analysis described below, promised franchise fees are significantly higher in renewal contracts than in concurrently-struck initial franchise awards, everything else being held constant.

[20] LOGHOMES represents the the log of the number of homes in a community passed by a system. Evidence presented in Zupan (1988) implies that the number of homes a system passes can be treated as an exogenous variable. LOGHOMES is predicted to have a positive effect on CHCAP, FFEE and COMCH. Statistical tests confirm the validity of the assumption that the number of homes passed has a nonlinear, diminishing effect on CHCAP, FFEE and COMCH.

[21] The variable INCOME is based on Census data (U.S. Bureau of the Census, 1983) and is predicted to have a positive effect on CHCAP, FFEE and COMCH.

[22] MUSTCARRY represents the number of over-the-air broadcast signals that are significantly viewed in a system's area and that, until 1985, were required by federal law to be carried by the local operator (cf., Multichannel News, September 16, 1985, p. 3). MUSTCARRY is predicted to have a positive effect on CHCAP, while its effect on FFEE and COMCH cannot be predicted a priori.

[23] YEAR represents the amount of time after January 1, 1980 that the renewal or initial contract was agreed upon (e.g., a renewal contract agreed to on February 15, 1981 would have an associated YEAR value of 1.13). YEAR is predicted to have a positive effect on CHCAP and LOCH if improvements in channel capacity technology continued to occur over the 1980 to 1984 time period (cf., Cablevision, December 7, 1981, p. 147). The effect of YEAR on FFEE cannot be predicted a priori.

[24] An attempt was made to account for other exogenous factors such as a community's density (miles of plant per number of homes passed) and the relative amount of multiple dwelling units in the community (multiple dwelling units are particularly likely to obtain their own satellite dishes and to not

rely on the franchised operator for cable services). Variables reflecting such factors, however, proved to be statistically insignificant when included in the Table 4 regression models.

[25] In cases of aerial wiring, there is no economic advantage to installing excess capacity (Malarkey, Taylor, and Associates, 1975). There may even be an economic disadvantage if changes in technology render the excess capacity obsolete. In 1985, for example, General Electric began marketing a signal compression technology which allows operators to double an existing cable's capacity without installing a new, larger cable or a second trunk (Cablevision, March 11, 1985, p. 14). Of the three dual cable, excess capacity systems in the initial franchise award sample, the second residential cable was entirely inactive at the time the system's operators were surveyed in the fall of 1984.

[26] DUALEXCESS reflects the number of channels in the second, inactive cable for the three dual cable systems in the initial franchise award sample. DUALEXCESS is predicted to have a positive effect on CHCAP, LOCH (if greater channel capacity implies an operator can dedicate more capacity to community programming), and FFEE (if a city's ability to obtain a significant amount of channel capacity also implies that the city will be able to negotiate a higher franchise fee). The Table 5 results suggest an interesting hypothesis: significant prior experience in dealing with a contractual partner may build up sufficient trust and working knowledge to prevent a city from requesting economically wasteful amounts of channel capacity; wasteful amounts of capacity only get contracted for where there is the absence of such a working relationship and trust.

[27] A cross-sectional comparison of the prices charged in October 1984 is employed since data on promised prices (at the time of franchising) were not collected from the surveyed initial award operators. Such a cross-sectional comparison is a valid test for operator opportunism at the time of refranchising provided that operators do not behave opportunistically during initial contract execution. The validity of such an assumption is supported by the evidence presented in Prager (1986) and Zupan (1987). Only the 23 renewal systems for which the Kagan newsletter provided more finely-detailed contractual data are included in the price comparison; for the other renewal cases accurate data are not available with regards to either the prices charged or whether or not these systems were subject to formal basic tier rate control in October 1984.

[28] EXPBPEN represents the expected (at the time of refranchising/franchising) ratio of basic tier subscribers to homes passed by cable and is predicted to have a positive effect on BPRICE, BPPCH and PPRICE. For the sample of 66 initial award systems, the expected basic penetration rate was provided by the surveyed system managers. For the sample of renewal systems, the expected basic penetration rate is assumed to be the existing penetration rate at the time of refranchising. All results of interest reported below remain unchanged if EXPBPEN is replaced by INCOME and MUSTCARRY.

[29] LOGHOMES is included as an explanatory variable to account for possible economies of scale (Eli Noam, 1985; and Bruce Owen and Peter Greenhalgh, 1983). LOGHOMES is predicted to have a negative effect on BPRICE, BPPCH and PPRICE. Attempts to account for additional predetermined cost factors such as a community's density or the franchise fee promised at franchising/

refranchising time produce results that differ in no significant way from those reported in Table 6.

[30] If the absence of basic rate control results in higher prices, UNREGBASIC should have a positive effect on (at least) BPRICE and BPPCH. The effect of UNREGBASIC on pay tier rates cannot be predicted a priori (according to federal law, pay tier rates may not be subject to rate control). On the one hand, UNREGBASIC should have a negative influence on pay rates since basic service is a perfect complement for pay service (i.e., one generally cannot get pay service without also purchasing the basic tier). On the other hand, UNREGBASIC should have a positive effect on pay rates to the extent that the absence of formal controls over basic rates indicates less concern for consumer interests (and informal price control) in general by local policymakers. Six of the 66 initial award systems and at least 3 of the 23 renewal systems were not subject to formal rate controls in October 1984.

[31] Two of initial franchise award systems are operated on a non-profit basis. NONPROFIT is predicted to have a negative effect on prices if franchisors fail to eliminate operator profits and if not-for-profit operators act so as to maximize cable subscribership. Both nonprofit operators stated that their primary goal was the promotion of basic subscribership. The emphasis on basic over pay subscribership is borne out by the regression results reported in Table 6 below.

[32] For example, systems with an expanded basic tier are predicted to have a lower-priced regular basic tier, ceteris paribus.

[33] All of the Table 6 results of interest remain unchanged when variables reflecting additional predetermined institutional factors are included in the model: variables reflecting the presence of an initial (operator-promised) freeze on pay tier rates or the absence of Home Box Office (HBO) as the lead pay tier service.

[34] The OLS results are rather remarkable given that there are two biases present in the data that should work towards finding prices to be higher in renewal systems than in initial award systems. First, the 3 (out of 23) Kagan renewal systems for which no information was provided as to whether the system was subject to formal rate control are all assumed to have been subject to such controls in October 1984. Second, for the Kagan renewal systems for which price data are not available for October 1984, the prices reported (in industry trade manuals) for early 1985 are assumed to be identical to the prices charged in October 1984. Appendix A provides the results that obtain when the renewal systems for which less finely-detailed data are available in the Kagan newsletter are added to the analysis. Since information is not available as to whether any of these additional systems were subject to rate controls in October 1984, all of the systems are assumed to have been rate controlled. In addition, where price data are not available for these additional renewal systems for October 1984 but are available for early 1985, the prices reported for early 1985 are assumed to be identical to the prices charged in October 1984.

APPENDIX A

BASIC AND LEAD PAY TIER PRICES: AN EXPANDED SAMPLE OF RENEWAL CONTRACTS

EXPLANATORY VARIABLE	BASIC PRICE	BASIC PRICE PER CHANNEL OF SERVICE	LEAD PAY PRICE
LOGHOMES	-0.04 (-0.39)	-0.10** (-6.88)	-0.16* (-1.78)
EXPBPEN	0.86 (0.81)	0.13 (0.89)	0.65 (0.68)
UNREGBASIC	2.80** (4.47)	0.17* (2.06)	0.18 (0.33)
NONPROFIT	-3.14** (-2.53)	-0.22 (-1.31)	-0.09 (-0.08)
EXPAND	-1.42** (-3.69)	0.17** (3.18)	0.28 (0.83)
RENEW	-0.19 (-0.55)	0.06 (1.29)	0.92** (2.98)
CONSTANT	9.20** (7.63)	1.29** (7.94)	10.57** (9.99)
RSQ	0.22	0.40	0.15
CRRSQ	0.18	0.37	0.10
TOTAL OBS.	115	115	112
RENEWAL OBS.	49	49	46

* Significant at the .05 level in a one-tailed test.
** Significant at the .01 level in a one-tailed test.

NOTE: T-statistics are in parentheses.

BIBLIOGRAPHY

Axelrod, Robert, The Evolution of Cooperation, New York: Basic Books, Inc., 1984.

Cable Television Information Center, CableReports, Washington, D.C.: CTIC, monthly, various issues.

Demsetz, Harold E., "Why Regulate Public Utilities?," Journal of Law and Economics, April 1968, Vol. 11, No. 1, pp. 55-65.

Dow Jones and Company, The Wall Street Journal, New York: Dow Jones and Company, daily.

Dubin, Jeffrey A. and Peter Navarro, "How Markets for Impure Public Goods Organize: The Case of Household Refuse Collection," unpublished Social Science Working Paper 633, California Institute of Technology, March 1987.

Edwards, Corwin D., "Conglomerate Bigness as a Source of Power," in the National Bureau of Economic Research conference report, Business Concentration and Price Policy, Princeton, New Jersey: Princeton University Press, 1955.

Fairchild Business Publications, Multichannel News, Denver: Fairchild Business Publications, weekly, various weeks.

Goldberg, Victor P., "Regulation and Administered Contracts," Bell Journal of Economics and Management Science, Autumn 1976, Vol. 7, No. 2, pp. 426-448.

International Thomson Communications, CableFile, Denver: International Thomson Communications, yearly, various years.

----------, CableVision, Denver: International Thomson Communications, weekly, various weeks.

Joskow, Paul L., "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation," Journal of Law and Economics, October 1974, Vol. 17, No. 2, pp. 291-327.

----------, "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants," Journal of Law, Economics, and Organization, Spring 1985, Vol. 1, No. 1, pp. 33-80.

----------, "Contract Duration and Relationship-Specific Investments: Empirical Evidence From Coal Markets," American Economic Review, March 1987, Vol. 77, No. 1, pp. 168-185.

---------- and Richard Schmalensee, "Incentive Regulation for Electric Utilities," Yale Journal on Regulation, Fall 1986, Vol. 4, No. 1, pp. 1-49.

Kagan Associates, Inc., *Cable TV Franchising*, Carmel, California: Paul Kagan Associates, Inc., bimonthly/weekly, various issues.

Klein, Benjamin, Robert Crawford, and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, October 1978, Vol. 21, No. 3, pp. 297-326.

---------- and Roy W. Kenney, "Contractual Flexibility," unpublished paper, revised January 10, 1986.

---------- and Keith Leffler, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, December 1981, Vol. 89, No. 4, pp. 615-641.

Laffont, Jean-Jacques and Jean Tirole, "Using Cost Observations to Regulate Firms," *Journal of Political Economy*, June 1986, Vol. 94, No. 3, Pt. 1, pp. 614-641.

Lewis, Tracy R., "Reputation and Contractual Performance in Long-Term Projects," *Rand Journal of Economics*, Summer 1986, Vol. 17 , No. 2, pp. 141-157.

Macauley, Stewart, "Non-contractual Relations in Business," *American Sociological Review*, February 1963, Vol. 28, No. 1, pp. 55-67.

Masten, Scott E., "The Organization of Production: Evidence from the Aerospace Industry," *Journal of Law and Economics*, October 1984, Vol. 27, No. 2, pp. 403-417.

Monteverde, Kirk and David J. Teece, "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics and Management Science*, Spring 1982, Vol. 13, No. 1, pp. 206-213.

Noam, Eli, "Economies of Scale in Cable Television: A Multi-Product Analysis," in Eli Noam, ed., *Video Media Competition*, pp. 93-120, New York: Columbia University Press, 1985.

Owen, Bruce M. and Peter R. Greenhalgh, "Competitive Policy Considerations in Cable Television Franchising," Washington, D.C.: Economists, Inc., 1983.

Palay, Thomas M., "Comparative Institutional Economics: The Governance of Freight Rail Contracting," *Journal of Legal Studies*, June 1984, Vol. 13, No. 2, pp. 265-288.

Peacock, Alan and Charles Rowley, "Welfare Economics and the Public Regulation of Natural Monopoly," *Journal of Public Economics*, Spring 1972, Vol. 3, No. 1, pp. 98-129.

Peltzman, Sam, "Toward a More General Theory of Regulation," *Journal of Law and Economics*, August 1976, Vol. 19, No. 2, pp. 211-240.

Prager, Robin A., "Firm Behavior in Franchise Monopoly Markets: The Case of Cable Television," unpublished Ph.D. dissertation, M.I.T., 1986.

Radner, Roy, "Competitive Equilibrium Under Uncertainty," *Econometrica*, January 1968, Vol. 36, No. 1, pp. 31-58.

Schelling, Thomas C., *The Strategy of Conflict*, Cambridge, Massachusetts: Harvard University Press, 1960.

Scherer, Frederic M., *Industrial Market Structure and Economic Performance*, 2nd ed., Boston: Houghton Mifflin Company, 1980.

Schmalensee, Richard, *The Control of Natural Monopolies*, Lexington, Massachusetts: D.C. Heath and Company, 1979.

Simon, Herbert A., "Theories of Decision-making in Economics and Behavioral Science," *American Economic Review*, June 1959, Vol. 49, No. 3, pp. 253-283.

----------, *Administrative Behavior*, 2nd ed., New York: The Macmillan Company, 1961.

----------, "Rationality as Product and Process of Thought," *American Economic Review*, May 1978, Vol. 68, No. 2, pp. 1-16.

Stocking, George W. and Myron W. Watkins *Cartels in Action: Case Studies in International Business Diplomacy*, New York: Twentieth Century Fund, 1946.

Television Digest, Inc., *Television and Cable Factbook*, Washington, D.C.: Television Digest, Inc., yearly, various years.

----------, *Weekly Cable Action Update*, Washington, D.C.: Television Digest, Inc., weekly, various weeks.

U.S. Bureau of the Census, *County and City Databook*, Washington, D.C.: U.S. Government Printing Office, 1983.

Williamson, Oliver, "Franchise Bidding for Natural Monopolies--in General and with Respect to CATV," *Bell Journal of Economics and Management Science*, Spring 1976, Vol. 7, No. 1, pp. 73-104.

Zupan, Mark A., "Reneging by Cable Operators on Their Franchise Promises: Opportunism or Economic Necessity?," unpublished paper, Department of Finance and Business Economics, USC School of Business, November 1987.

----------, "Franchise Bidding for Natural Monopolies: How well does it Promote Efficiency in Local CATV Markets?," unpublished paper, Department of Finance and Business Economics, USC School of Business, January 1988.